#6

PATENT
Attorney Docket No.: 16869S-029800US
Client Ref. No.: E6042-01EW

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

RECEIVED
MAR 1 1 2004
Technology Center 2100

In re application of:

Yoshihito Nakagawa et al.

Application No.: 09/915,692

Filed: July 25, 2001

For: Storage-Related Accounting System and Method of the Same

Customer No.: 20350

Examiner: H. Alam

Technology Center/Art Unit: 2152

PETITION TO MAKE SPECIAL FOR NEW APPLICATION PURSUANT TO 37 C.F.R. § 1.102(d) & M.P.E.P. § 708.02, Item VIII, ACCELERATED EXAMINATION

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

This is a petition to make special the above-identified application under MPEP § 708.02, Item VIII, accelerated examination. The application has not received any examination by the Examiner.

(A) The Commissioner is authorized to charge the petition fee of $130 under 37 C.F.R. § 1.17(h), and any additional fees that may be associated with this petition may be charged to Deposit Account No. 20-1430.

03/10/2004 HALI11 00000030 201430 09915692
01 FC:1460 130.00 DA

(B) All the claims are believed to be directed to a single invention. If the examiner determines that all the claims presented are not obviously directed to a single invention, then Applicant will make an election without traverse as a prerequisite to the grant of special status where the specific grouping of claims will be determined by the examiner.

(C) A pre-examination search was made by a foreign patent office in connection with a corresponding foreign application. The corresponding foreign application is European Patent application No. EP01117862, a copy of the first page of which is provided herewith for identification purposes as Exhibit A. A search was performed by an examiner in the European Patent Office for the European patent application, and reported in a communication dated October 29, 2003 ("EPO search report"), submitted herewith as Exhibit B. The following references were identified in the EPO search report:

(1) U.S. Pat. No. 6,119,160 to Zhang et al.;

(2) Rigney, C., "RFC 2866: RADIUS Accounting" (June 2000);

(3) Mills, C., "RFC1272: INTERNET Accounting: Background" (November 1991); and

(4) PCT International Publication No. WO 97/22936.

(D) The references cited in the EPO search report have been made of record in an Information Disclosure Statement mailed December 8, 2003. Copies of the references are enclosed herewith for convenience, collectively identified as Exhibit C.

(E) Set forth below is a detailed discussion of the references, pointing out with particularly how the claimed subject matter, recited in the claims as amended per the preliminary amendment filed herewith, is distinguishable over the references.

Claimed Subject Matter of the Present Invention

The present invention is directed to an accounting system for data storage systems. Briefly, the present invention provides accounting based on the number of times an accounting object is accessed, and on the amount of data transfer that occurs with the accounting object.

Claim 1 recites an accounting system including host computers, an accounting server, a storage control device, and storage devices for storing data. The accounting system

comprises "accounting data generating means for generating accounting data comprising a number of times of access and a data transfer quantity for each of one or more accounting subject control units." The accounting system recited in claim 1 further comprises a "transfer means for informing said accounting server of said accounting data generated by said accounting data generating means." Claim 2 recites that the accounting subject control units "includes at least one of a host computer, a World Wide Name, a channel port, a storage device and an in-storage-device area."

A further aspect of the invention is the storage control device "is associated with a service processor and includes transfer means for informing said service processor of said accounting data" as recited in Claim 3. Still a further aspect of the invention includes "means for setting an upper limit of said accounting data ... made to operate from at least one of said accounting server and said service processor" as recited in Claim 4. The upper limit value can be "determined on the basis of a predetermined period which includes, as a unit, at least one of a day, a week, a month, and a year" as recited in Claim 5. The storage control device can include means which "does not accept said data input/output request larger than said upper limit value" as recited in Claim 6.

Claim 12 recites a method for making accounting for a plurality of host computers in a storage-related accounting system. Recited steps include "designating accounting subject control units." Claim 13 recites that accounting subject control units "include host computers, World Wide Names, channel ports, and storage areas in storage devices." The method recited in Claim 12 further includes "for at least some of said accounting subject control units, determining an upper limit value for a number of times of access and for an upper limit value for a data transfer quantity" and "generating accounting data of said accounting subject control units in accordance with said upper limit values."

U.S. Patent No. 6,119,160 to Zhang et al.

The reference to Zhang et al. relates to time-based and byte-based accounting for start/stop events. *Col. 1, lines 5 - 12.* Events include the user account logon, service establishments, and the Point to Point protocol (PPP) connections within the network. A counter

tracks the duration of sessions and connections and the byte-count associated with the specified session or connection. *Col. 2, lines 20 - 30.*

Figs 2A and 2B constitute a flow chart of a typical logon-to-logoff session, showing how accounting is performed. The description is given from col. 3, line 64 to col. 8, line 20. Highlights of the flow chart include the accounting events such as logon, service logon (establishment), service connection, and the respective reverse activities (e.g., logoff). Accounting for a logon event begins at step 46, and is described beginning at about col. 4, line 36. Accounting for a service logon event begins at step 72, and is described beginning at about col. 5, line 20. Accounting for a service PPP connection event begins at step 78, and is described beginning at about col. 5, line 44.

Accounting for a disconnect from a PPP connection begins at step 84, and is described at about col. 6, line 10. Attributes associated with this accounting event are shown in col. 6, lines 23 - 53. Of note is the Acct-Session-Time which is the length of the session in seconds.

Accounting for the termination of a service begins at step 90, and is described at about col. 6., line 64. The attributes for this event are shown in col. 7, lines 10 and following. Of note is the Acct-Session-Time which measures the session duration in seconds.

Finally, accounting for logoff begins at step 96, and is described at about col. 7, line 32.

Zhang et al. do not appear to show the subject matter recited in the pending claims. They do not teach "accounting data comprising a number of times of access and a data transfer quantity for each of one or more accounting subject control units." Claim 1. Zhang et al. clearly disclose time-based accounting, but do not show number-of-times-of-access accounting. Moreover, the time-based attributes for the logoff and disconnection events do not suggest accounting based on a number of times of access to an accounting subject unit.

Zhang et al. do not teach "determining an upper limit value for a number of times of access and for an upper limit value for a data transfer quantity" of an accounting subject unit, nor do they teach "generating accounting data of said accounting subject control units in accordance with said upper limit values." Claim 12. First, Zhang et al. teach to simply monitor

the time duration of a session; they do not teach placing an upper limit to a session. Secondly, monitoring a duration of time does not suggest monitoring the number of times of access to an accounting subject unit.

RFC 2866: RADIUS Accounting by Rigney, C.

This reference describes a protocol for carrying out accounting information between a network access server and a shared accounting server. The protocol is known as RADIUS (remote authentication dial-in user service). This reference discloses an extension of the RADIUS protocol to cover delivery of accounting information from the network access server to a RADIUS accounting server. *Page 2.*

Terminology used in the document is discussed on page 3. Of note is the term "session" of a service, characterized by a beginning-of-session and an end-of-session. A user may have multiple sessions in parallel or in series, with each session generating a start and a stop accounting record with its own session ID.

An Accounting Start packet is transmitted at the start of a service delivery. Conversely, an Accounting Stop packet is sent at the end of service delivery. *Page 4.*

Packet format is described on pages 5 - 7. Packet types are discussed on pages 7 - 9. A packet is either an Accounting-Request (page 8), or an Accounting-Response (page 9).

As shown on page 6, a packet may have attributes. Attributes carry the specific details for the two packet types. *Page 10.* The attributes described in this reference are shown on page 11.

A noteworthy attribute is the Acct-Session-Time attribute discussed on page 17. This attribute indicates how many seconds the user has received a service.

The attribute Acct-Multi-Session-Id discussed on page 21 make it easy to link together multiple related sessions. Each session linked together would have a unique Acct-Session-Id (see page 15), but the same Acct-Multi-Session-Id.

Pages 23 - 25 show a list of attributes which may be found in Accounting-Request type packets.

This reference does not appear to show the subject matter recited in the pending claims. It does not teach "accounting data comprising a number of times of access and a data transfer quantity for each of one or more accounting subject control units." Claim 1. The Acct-Session-Time attribute does not measure number of times of access. None of the other attributes listed on page 11 and on pages 23 - 25 show or suggest an attribute that measures number of times of access.

This reference does not appear to disclose "determining an upper limit value for a number of times of access and for an upper limit value for a data transfer quantity" of an accounting subject unit, nor do they teach "generating accounting data of said accounting subject control units in accordance with said upper limit values." Claim 12. A review of the attributes does not show any attributes for setting a limit for number of times of access. A Port-Limit attribute is identified on page 25, but is not otherwise described.

Portions of the reference appear to have been identified by the examiner who prepared the EPO search report, by underlining text and annotating the margins. The identified portions include pages 2 - 4, 6, 8 - 10, 13, and 26. A review of these identified portions does not reveal any teaching of the subject matter recited in the claims.

RFC 1272: INTERNET Accounting: Background by Mills et al.

This reference provides background information for the "Internet Accounting Architecture." The focus is on defining METER SERVICES and USAGE REPORTING. A goal of a usage reporting architecture is to define a generalized accounting management activity, including calculations, usage reporting to users and providers and enforcing various limits on the use of resources. The authors discuss the motivation for usage reporting, on pages 2 and 3.

Accounting requirements are driven by policy, and vice-versa where policy is influenced by available management and reporting tools. *Section 3.2, pages 4 - 6.* The authors note that existing accounting policies for network traffic are usually based on parameters which change seldom; e.g., line speed of a physical connection which leads to FLAT-FEE billing. An incentive for this type of billing includes predictable monthly charges. Another inventive is no overhead is incurred for managing or tracking packet counts and usage-based reports. The

authors further note that USAGE-SENSITIVE charges may be preferred, however, by low-volume users.

In Section 3.3 (pages 6 - 8), the authors discuss some underlying assumptions they make to simplify the environment for developing an architecture for internet usage accounting.

A METER is a process which examines packets on a communications medium, and records aggregate counts of packets. *Page 8.* Meter placement is discussed on pages 9 - 10. Meter types include network monitors, line monitors, router-integral monitors, and router spiders. *Page 10.* The GRANULARITY of the information collected by a meter is controlled by the following considerations, described on pages 11 - 12. The ENTITY that is metered affects granularity. A PORT entity in the network is the coarsest granularity, while a HOST entity and a USER entity are at the fine-end of the granularity scale. The ATTRIBUTE categorizes packets into types; at the finest granularity, packets are individually identified by type of service being communicated. Other factors include: VALUES with units such as simple counters (packet, byte) and start/stop time; and REPORTING INTERVAL. Ways of collecting metering information are discussed in Section 4.4, pages 12 - 14.

A series of examples to illustrate the kinds of data that might be of interest to service providers and consumers is described in Section 5, on pages 14 - 17. In a single LAN service (Section 5.1), flows between individual host pairs can be measured. In an extended LAN (Section 5.2), an administrator might collect flow information between subnets in the LAN, while subnets might want to keep track of flows between hosts. Section 5.3 and Section 5.4 discuss typical data of interest at the regional and national backbone levels, respectively.

The authors provide a brief discussion about future issues: the need for application standards for accounting systems; quotas; session accounting, which refers to detailed auditing of individual sessions across the internet; application level accounting, which refers to hosts and proxy agents performing their own accounting; and the user.

From the foregoing summary of the Mills et al. reference, it appears there is no discussion as to elements of the present invention as recited in the pending claims. The reference does not teach "accounting data comprising a number of times of access and a data transfer

quantity for each of one or more accounting subject control units." Claim 1. The reference does not appear to disclose "determining an upper limit value for a number of times of access and for an upper limit value for a data transfer quantity" of an accounting subject unit, nor do they teach "generating accounting data of said accounting subject control units in accordance with said upper limit values." Claim 12.

PCT Published Application: WO 97 22936 to Eggleston et al.

Eggleston et al. describe a system for monitoring and controlling the amount of communications between a remote unit and a communication server. Five embodiments are disclosed. *Pages 4 - 6.* A virtual session manager is provided, in a first embodiment, to establish and maintain a session-less communication path with a first device and a session-oriented path with a second device. In a second embodiment, a pre-stage filter is provided to apply user-defined filter parameters on data being transferred between a remote unit and a server unit. In a third, embodiment, a select and summary listing is used to allow a user to review and request filtered data. In a fourth embodiment, an optimized technique for replying to messaging is provided. In a fifth embodiment, a rate governor monitors and controls the amount of communications between remote and server units.

Eggleston et al. disclose the use of thresholds in the rate governor to alert and restrict communications. Kindly refer to pages 24 and following. The rate governor tracks the approximate time and/or expense for client use. This includes the timing of a circuit-switched connection, or timing or estimating size of transmitted packets. Time or cost thresholds can be used to trigger alerts to warn the user of remaining time or charge.

A review of Eggleston et al. does not appear to show the subject matter recited in the pending claims. Eggleston et al. do not teach "accounting data comprising a number of times of access and a data transfer quantity for each of one or more accounting subject control units." Claim 1.

Eggleston et al. do not appear to disclose "determining an upper limit value for a number of times of access and for an upper limit value for a data transfer quantity" of an

accounting subject unit, nor do they teach "generating accounting data of said accounting subject control units in accordance with said upper limit values." Claim 12.

In view of this petition, the Examiner is respectfully requested to issue a first Office Action at an early date.

Respectfully submitted,

George B. F. Yee
Reg. No. 37,478

TOWNSEND and TOWNSEND and CREW LLP
Two Embarcadero Center, 8th Floor
San Francisco, California 94111-3834
Tel: 650-326-2400
Fax: 415-576-0300
Attachments
GBFY:cmm
60134267 v1

Exhibit A

"Corresponding European Patent Application, Title Page"

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 235 384 A2

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
28.08.2002 Bulletin 2002/35

(51) Int Cl.[7]: **H04L 12/24**

(21) Application number: **01117862.1**

(22) Date of filing: **23.07.2001**

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE TR
Designated Extension States:
AL LT LV MK RO SI

(30) Priority: **08.02.2001 JP 2001031707**

(71) Applicant: **Hitachi, Ltd.**
Chiyoda-ku, Tokyo 101-8010 (JP)

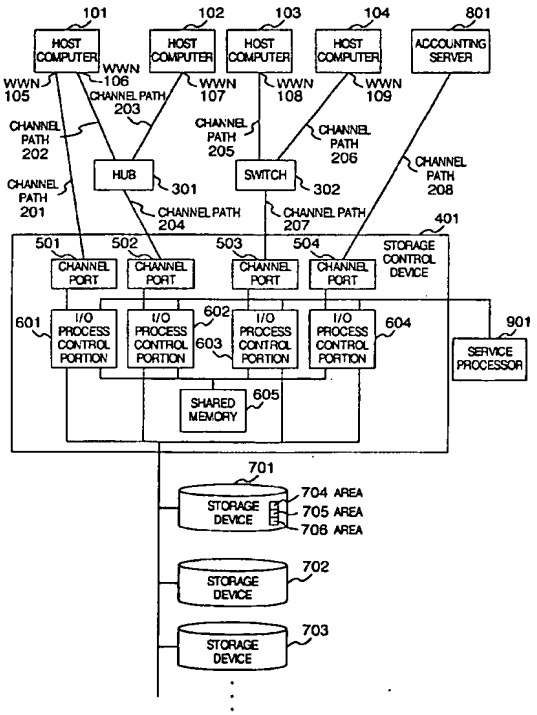
(72) Inventors:
- **Nakagawa, Yoshihito**
  Chiyoda-ku, Tokyo 100-8220 (JP)
- **Yokohata, Shizuo**
  Chiyoda-ku, Tokyo 100-8220 (JP)

(74) Representative: **Strehl Schübel-Hopf & Partner**
Maximilianstrasse 54
80538 München (DE)

(54) Accounting system and method for storage devices

(57) An accounting system includes a plurality of host computers (101 to 104), an accounting server (801) connected to the plurality of host computers, a storage control device (401) connected to the plurality of host computers and the accounting server and having a plurality of channel ports (501 to 504) for performing data input/output operation, storage devices (701 to 703) connected to the storage control device for storing data inputted/outputted to/from the plurality of host computers, an accounting data generating unit (1006) for generating accounting data containing at least one of number of times of access and a data transfer quantity for every accounting subject control unit, and a transfer unit (1009) for informing the accounting server (801) of the accounting data generated by the accounting data generating unit, executing the accounting process for the plurality of accounting subject control units.

FIG.1



EP 1 235 384 A2

Exhibit B

"European Search Report"

P.B.5818 – Patentlaan 2
2280 HV Rijswijk (ZH)
+31 70 340 2040
TX 31651 epo nl
FAX +31 70 340 3016

Europäisches Patentamt
Zweigstelle in Den Haag Recherchen-abteilung

European Patent Office
Branch at The Hague Search division

Office européen des brevets
Département à La Haye Division de la recherche

Strehl Schübel-Hopf & Partner
Maximilianstrasse 54
80538 München
ALLEMAGNE

Erhalten
29.OKT.2003
Strehl et al.

Datum/Date
29.10.03

Zeichen/Ref./Réf.	Anmeldung Nr./Application No./Demande n°./Patent Nr./Patent No./Brevet n°.
EPA-38968	01117862.1-2413-

Anmelder/Applicant/Demandeur/Patentinhaber/Proprietor/Titulaire
Hitachi, Ltd.

COMMUNICATION

The European Patent Office herewith transmits as an enclosure the European search report for the above-mentioned European patent application.

If applicable, copies of the documents cited in the European search report are attached.

☐ Additional set(s) of copies of the documents cited in the European search report is (are) enclosed as well.

The following specifications given by the applicant have been approved by the Search Division:

☒ abstract ☒ title

☐ The abstract was modified by the Search Division and the definitive text is attached to this communication.

The following figure will be published together with the abstract: 1

EUROPÄISCHES PATENTAMT - DEN HAAG - EUROPEAN PATENT OFFICE - THE HAGUE - OFFICE EUROPÉEN DES BREVETS - LA HAYE

REFUND OF THE SEARCH FEE

If applicable under Article 10 Rules relating to fees, a separate communication from the Receiving Section on the refund of the search fee will be sent later.

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 01 11 7862

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.7)
X	US 6 119 160 A (SITARAMAN ARAVIND ET AL) 12 September 2000 (2000-09-12) * column 3, line 31 - column 9, line 48 *	1,3	H04L12/24 H04L12/14
Y		2,4-17	
X	RIGNEY C: "RFC2866: RADIUS Accounting" INTERNET DRAFT, June 2000 (2000-06), XP002205198 Retrieved from the Internet: <URL:www.ietf.org/rfc/rfc2866.txt> 'retrieved on 2002-06-28! section "Operation": * page 4 * Section 4 "Packet Types": * page 7 - page 10 * Sections 5.3 - 5.5 of Section 5 "Attributes" * page 14 - page 16 *	1,3	
A		2,4-17	
Y	C. MILLS, D. HIRSH, & G. RUTH : "INTERNET ACCOUNTING: BACKGROUND" INTERNET DRAFT, 'Online! XP002255773 Retrieved from the Internet: <URL:http://www.watersprings.org/pub/rfc/rfc1272.txt> 'retrieved on 2003-09-25! Sections 2 "Goals for a Usage Reporting Architecture" and sections 3.1 - 3.2 of section 3 "The Usage Reporting Function": * page 2 - page 6 *	2,4-17	TECHNICAL FIELDS SEARCHED (Int.Cl.7) H04L
A	WO 97 22936 A (MOTOROLA INC) 26 June 1997 (1997-06-26) * abstract * * page 24, line 12 - page 27, line 2 *	2,4-17	

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
MUNICH	1 October 2003	Ross, C

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

ANNEX TO THE EUROPEAN SEARCH REPORT ON EUROPEAN PATENT APPLICATION NO. EP 01 11 7862

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report. The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

01-10-2003

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
US 6119160	A	12-09-2000	NONE		
WO 9722936	A	26-06-1997	CA	2216533 A1	26-06-1997
			CA	2365520 A1	26-06-1997
			GB	2314729 A ,B	07-01-1998
			US	2002013854 A1	31-01-2002
			WO	9722936 A1	26-06-1997
			US	2003084184 A1	01-05-2003

Exhibit C

"References (4) cited in a European Search Report"

US006119160A

United States Patent [19]

Zhang et al.

[11] Patent Number: 6,119,160

[45] Date of Patent: Sep. 12, 2000

[54] MULTIPLE-LEVEL INTERNET PROTOCOL ACCOUNTING

[75] Inventors: **Shujin Zhang**, San Mateo; **Shuxian Lou**, San Jose; **Roman Peter Kochan**, Irvine; **Aravind Sitaraman**, Santa Clara, all of Calif.

[73] Assignee: **Cisco Technology, Inc.**, San Jose, Calif.

[21] Appl. No.: **09/172,183**

[22] Filed: **Oct. 13, 1998**

[51] **Int. Cl.**[7] **G06F 15/173**

[52] **U.S. Cl.** **709/224**; 709/229

[58] **Field of Search** 379/127; 705/14, 705/30; 709/223, 224, 217, 218, 219, 225, 229

[56] **References Cited**

U.S. PATENT DOCUMENTS

5,283,783	2/1994	Nguyen et al.	370/222
5,287,103	2/1994	Kasprzyk et al.	370/401
5,519,704	5/1996	Farinacci et al.	370/402
5,555,244	9/1996	Gupta et al.	370/397
5,592,470	1/1997	Rudrapatna et al.	370/320
5,621,721	4/1997	Vatuone .	
5,668,857	9/1997	McHale	379/93.07
5,673,265	9/1997	Gupta et al.	370/432
5,678,006	10/1997	Valizadeh et al.	709/223
5,717,604	2/1998	Wiggins	709/229
5,729,546	3/1998	Gupta et al.	370/434
5,740,176	4/1998	Gupta et al.	370/440
5,745,556	4/1998	Ronen	379/127
5,768,521	6/1998	Dedrick	709/224
5,778,182	7/1998	Cathey et al.	709/219
5,787,253	7/1998	McCreery et al.	709/231
5,799,017	8/1998	Gupta et al.	370/419
5,815,665	9/1998	Teper et al.	709/229
5,852,812	12/1998	Reeder	705/39
5,870,550	2/1999	Wesinger, Jr. et al.	709/218
5,898,780	4/1999	Liu et al.	380/25

(List continued on next page.)

OTHER PUBLICATIONS

Active Software, Inc., "Active Software's Integration System", printed from http://www.activesw.com/products/products.html, on Jul. 24, 1998, 25 pages.

Cisco Systems, Inc., "CiscoDNS/DHCP Manager V.1.1", printed from http://www.combinet.com/warp/public/751/dnsmg/dnsmg_ds.htm, on Sep. 10, 1998, 4 pages.

Cisco Systems, Inc., "Cisco DNS/DHCP Manager V.1.1", printed from http://www.combinet.com/warp/public/751/dnsmg/dnsmg_pa.htm, on Sep. 10, 1998, 7 pages.

Cisco Systems, Inc., "DHCP Solution Helps Scale and Configure IP Nodes in Growing Switched Networks", printed from http://cio.cisco.co.jp/warp/public/795/6.html, on Sep. 10, 1998, 2 pages.

Cisco Systems, Inc., "Cisco DNS/DHCP Manager", printed from http://mwrns.noaa.gov/cisco/cc/td/doc/resprdct/res31.htm, on Sep. 10, 1998, 4 pages.

Edell, et al., "Billing Users and Pricing for TCP", 1995, IEEE Journal on Selected Areas in Communications, pp. 1–14.

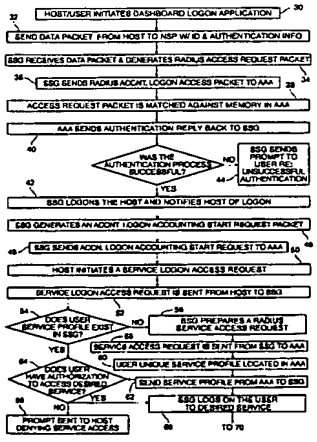
Network Registrar, "Hot Products & Solutions—IP Address Management: A White Paper", American Internet Corporation, Bedford, MA, printed from http://www.american.com/ip–mgmt.html, on Jul. 24, 1998.

Primary Examiner—Ahmad F. Matar
Assistant Examiner—Patrice Winder
Attorney, Agent, or Firm—D'Alessandro & Ritchie

[57] **ABSTRACT**

A method and apparatus for providing computer network access points the capability for multiple-level accounting. A gateway device located at the access point is capable of generating Internet protocol accounting start and stop requests based on various events that need to be accounted for when a user accesses a network. These events include the user account logon, the service establishments and the Point to Point protocol (PPP) connections between the gateway device and public and private domains within the network. The counter is capable of tracking the duration of sessions and connections and the byte-count associated with the specified session or connection. The gateway device communicates with an accounting server which stores the accounting requests and matches start requests with subsequent stop requests.

22 Claims, 4 Drawing Sheets



U.S. PATENT DOCUMENTS

5,913,037	6/1999	Dpofford et al.	709/226
5,918,016	6/1999	Brewer et al.	709/220
5,960,409	9/1999	Wexler	709/224 X
5,970,477	10/1999	Roden	709/229
5,992,051	7/1999	Sidey	709/223
6,018,619	1/2000	Allard et al.	709/224
6,026,440	2/2000	Shrader et al.	709/224
6,035,281	3/2000	Crosskey et al.	705/14

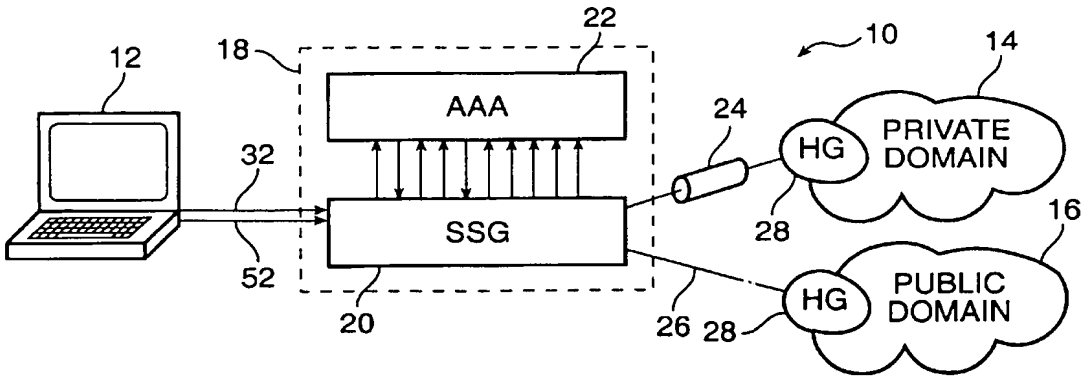

12
18
22
10
14
AAA
24
HG
PRIVATE DOMAIN
32
SSG
28
16
52
HG
PUBLIC DOMAIN
20
26
28

FIG. 1

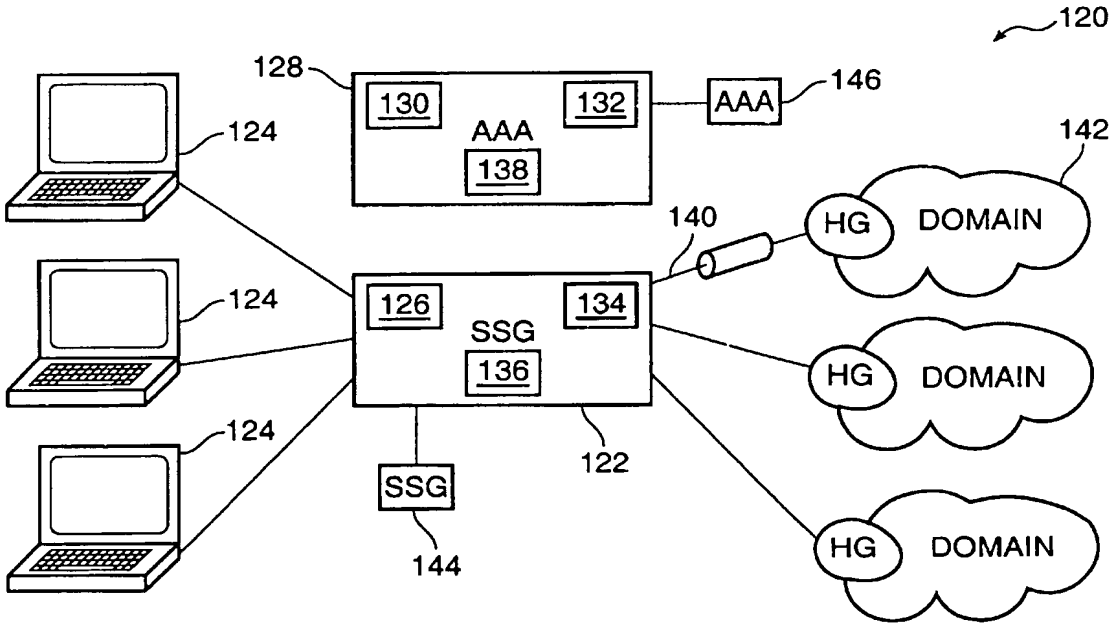

120
128
130
AAA
132
138
AAA
146
124
142
140
HG
DOMAIN
124
126
SSG
134
136
HG
DOMAIN
124
122
SSG
144
HG
DOMAIN

FIG. 3

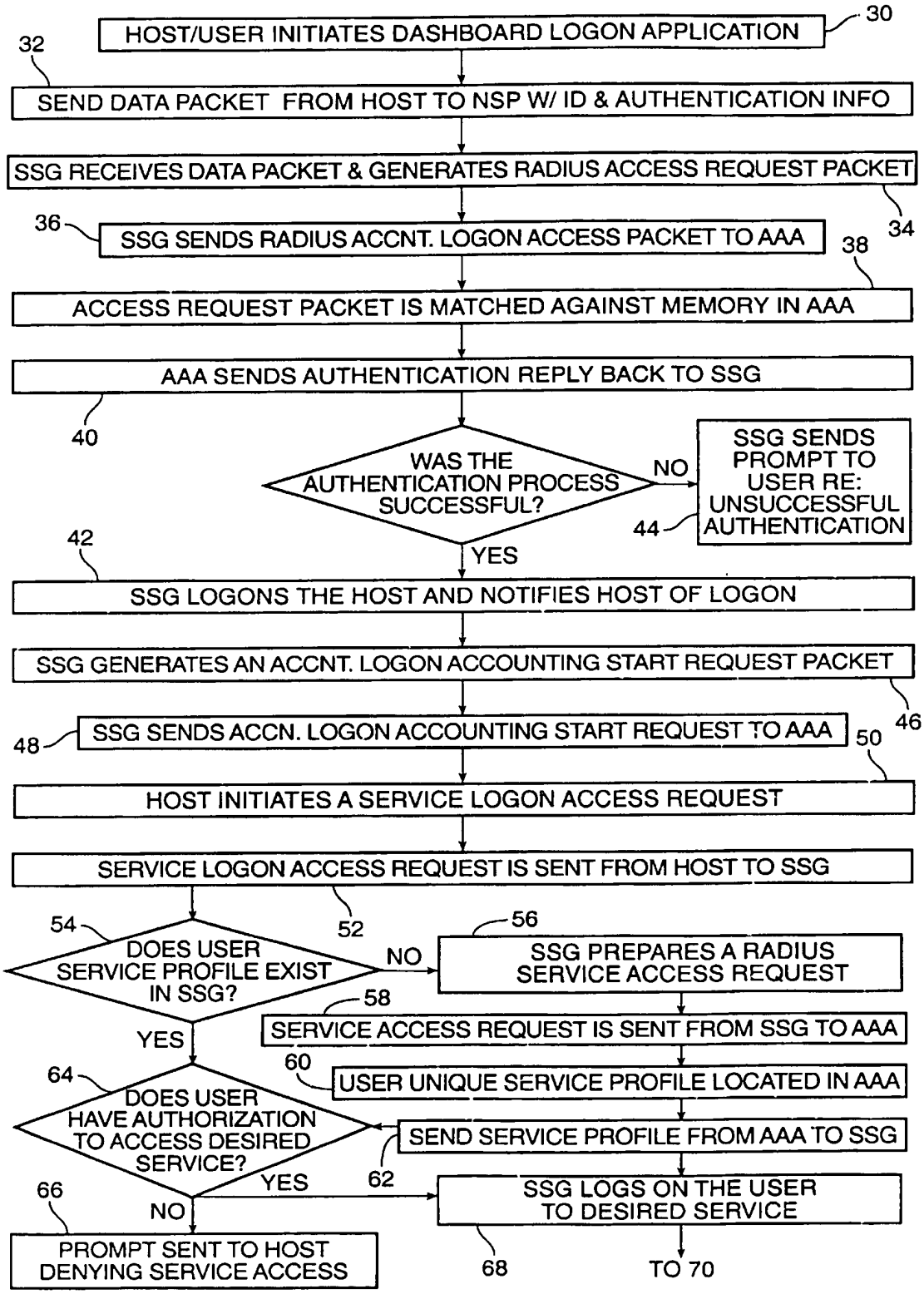

HOST/USER INITIATES DASHBOARD LOGON APPLICATION
30
32
SEND DATA PACKET FROM HOST TO NSP W/ ID & AUTHENTICATION INFO
SSG RECEIVES DATA PACKET & GENERATES RADIUS ACCESS REQUEST PACKET
34
36
SSG SENDS RADIUS ACCNT. LOGON ACCESS PACKET TO AAA
38
ACCESS REQUEST PACKET IS MATCHED AGAINST MEMORY IN AAA
AAA SENDS AUTHENTICATION REPLY BACK TO SSG
40
WAS THE AUTHENTICATION PROCESS SUCCESSFUL?
NO
SSG SENDS PROMPT TO USER RE: UNSUCCESSFUL AUTHENTICATION
44
YES
42
SSG LOGONS THE HOST AND NOTIFIES HOST OF LOGON
SSG GENERATES AN ACCNT. LOGON ACCOUNTING START REQUEST PACKET
46
48
SSG SENDS ACCN. LOGON ACCOUNTING START REQUEST TO AAA
50
HOST INITIATES A SERVICE LOGON ACCESS REQUEST
SERVICE LOGON ACCESS REQUEST IS SENT FROM HOST TO SSG
52
54
DOES USER SERVICE PROFILE EXIST IN SSG?
NO
56
SSG PREPARES A RADIUS SERVICE ACCESS REQUEST
58
SERVICE ACCESS REQUEST IS SENT FROM SSG TO AAA
YES
60
USER UNIQUE SERVICE PROFILE LOCATED IN AAA
64
DOES USER HAVE AUTHORIZATION TO ACCESS DESIRED SERVICE?
SEND SERVICE PROFILE FROM AAA TO SSG
62
YES
66
NO
SSG LOGS ON THE USER TO DESIRED SERVICE
PROMPT SENT TO HOST DENYING SERVICE ACCESS
68
TO 70

FIG. 2A

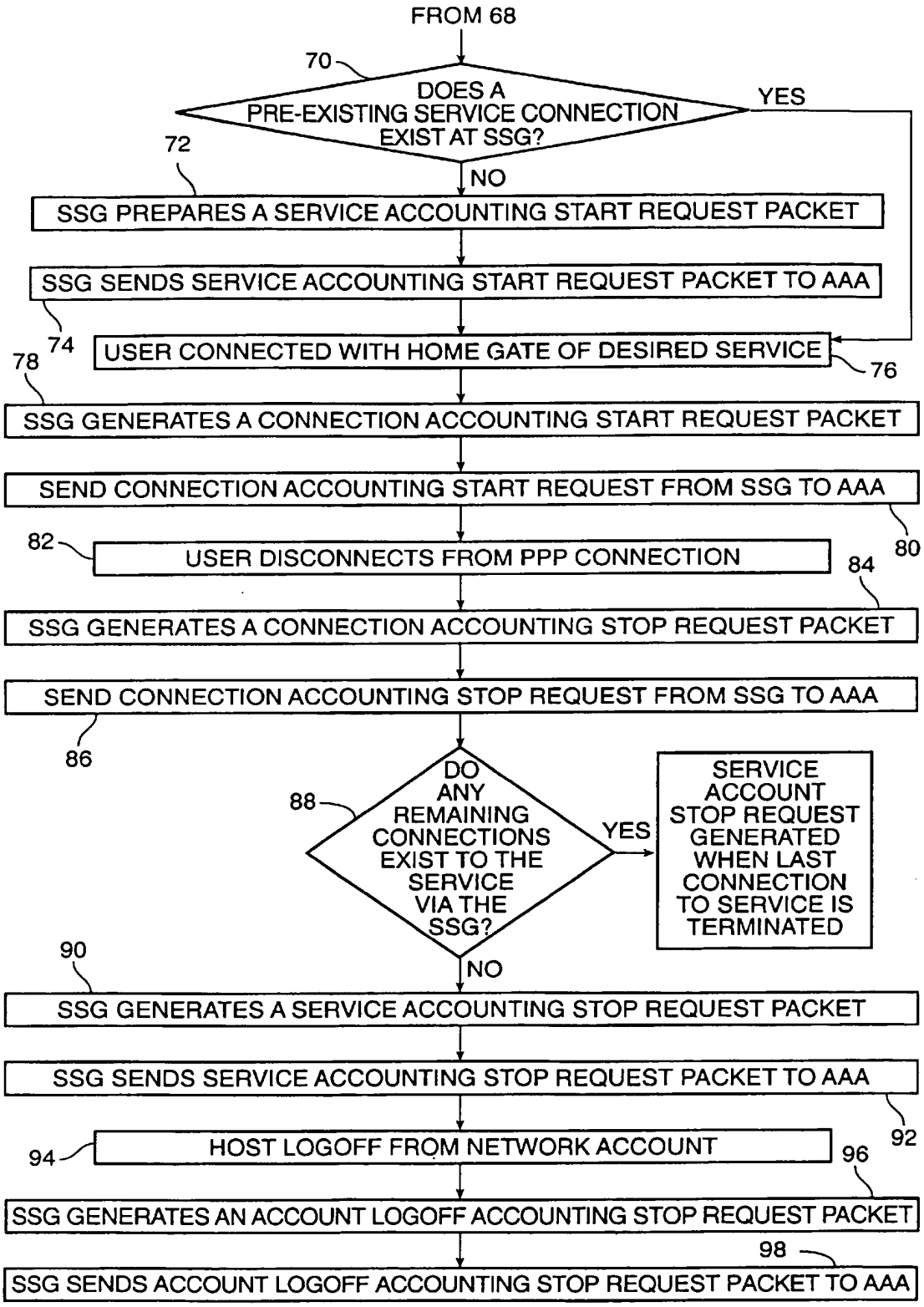

FROM 68
70
DOES A PRE-EXISTING SERVICE CONNECTION EXIST AT SSG?
YES
NO
72
SSG PREPARES A SERVICE ACCOUNTING START REQUEST PACKET
SSG SENDS SERVICE ACCOUNTING START REQUEST PACKET TO AAA
74
USER CONNECTED WITH HOME GATE OF DESIRED SERVICE
76
78
SSG GENERATES A CONNECTION ACCOUNTING START REQUEST PACKET
SEND CONNECTION ACCOUNTING START REQUEST FROM SSG TO AAA
80
82
USER DISCONNECTS FROM PPP CONNECTION
84
SSG GENERATES A CONNECTION ACCOUNTING STOP REQUEST PACKET
SEND CONNECTION ACCOUNTING STOP REQUEST FROM SSG TO AAA
86
88
DO ANY REMAINING CONNECTIONS EXIST TO THE SERVICE VIA THE SSG?
YES
SERVICE ACCOUNT STOP REQUEST GENERATED WHEN LAST CONNECTION TO SERVICE IS TERMINATED
NO
90
SSG GENERATES A SERVICE ACCOUNTING STOP REQUEST PACKET
SSG SENDS SERVICE ACCOUNTING STOP REQUEST PACKET TO AAA
92
94
HOST LOGOFF FROM NETWORK ACCOUNT
96
SSG GENERATES AN ACCOUNT LOGOFF ACCOUNTING STOP REQUEST PACKET
98
SSG SENDS ACCOUNT LOGOFF ACCOUNTING STOP REQUEST PACKET TO AAA

FIG. 2B

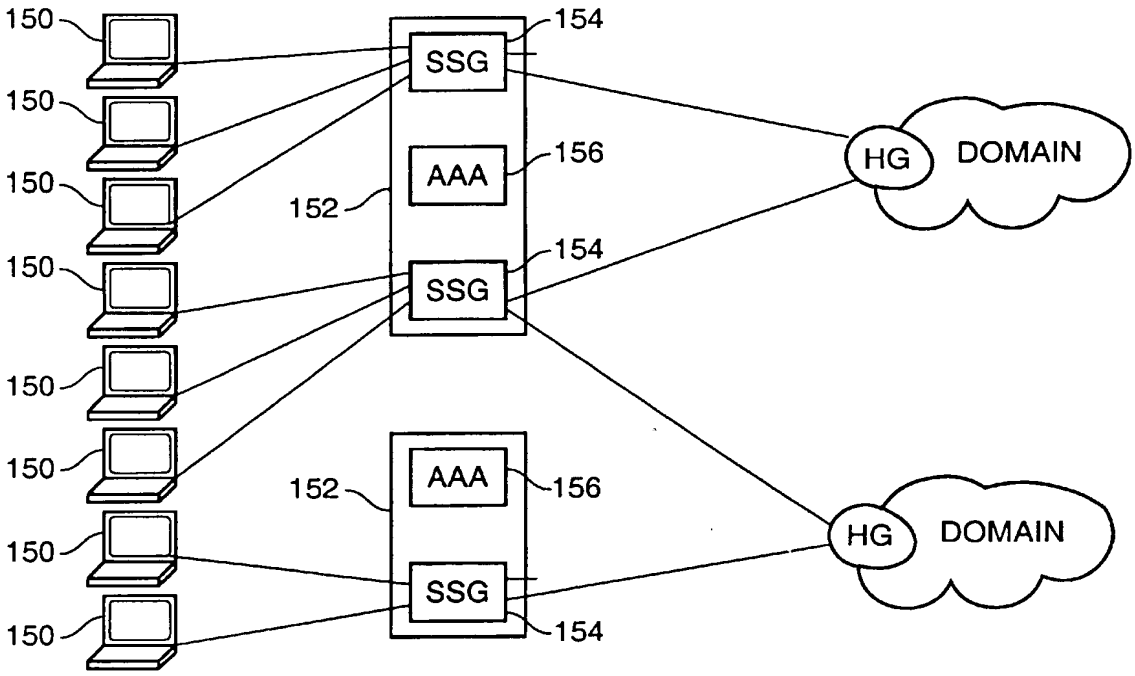

150
150
150
150
150
150
150
150
SSG
AAA
SSG
AAA
SSG
152
152
154
156
154
156
154
HG
DOMAIN
HG
DOMAIN

FIG. 4

MULTIPLE-LEVEL INTERNET PROTOCOL ACCOUNTING

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates to an accounting method and apparatus used in a computer network. More particularly, the present invention relates to a method and apparatus for generating time-based and/or byte-based accounting for various significant start/stop events using standard Internet protocol, such as the Remote Authentication Dial-In User Service (RADIUS) protocol.

2. Background

The ability to provide computer networking capabilities to the home personal computer (PC) is typically provided by telephone companies (Telcos) or commercial Internet Service Providers (ISPs) who operate network access points along the information superhighway. Network access points which are commonly referred to as Points of Presence or PoPs are located within wide area networks (WAN) and serve to house the network interfaces and service components necessary to provide routing, bridging and other essential networking functions. It is through these network access points that the user is able to connect with public domains, such as the Internet and private domains, such as the user's employer's intra-net.

Currently, Telcos and ISPs are limited in the means by which they can charge customers for their product. Basically, Telcos and ISPs are confined to either charging a flat fee, typically on a monthly basis, thus allowing the user unlimited network access for the specified period, or charging the user on a rate basis, typically an hourly rate. These billing schemes are primitive because the current capabilities possessed by the Telcos and ISPs provide only a simplified means of accounting for the events which a user undertakes during the time the user is logged on to the access point. Current technology only allows for the Telco or ISP to account for the duration of the period from when a user logs-on to the Telco or ISP and when the user subsequently logs-off.

As an example, a user implements a "dashboard" application on their host/computer which requires them to input identification and authorization information. This information is then sent via modem and telephone line to the Telco or ISP operated access point. A network access server (NAS) receives the identification and authorization information and proxies it to an authentication, authorization and accounting server. Once the server verifies the user authentication and authorization it grants the user logon access to downstream public and private networks. At this point a counter within the NAS is engaged which begins tracking the duration of the log-on session as well as the byte count encountered during the session. Subsequently, when the user desires to log off or a log off is warranted by other means outside of the control of the user, the counter within the NAS is disengaged and the appropriate accounting data is forwarded to the accounting server.

The Telco or ISP would benefit from having a more developed accounting scheme which allows for the tracking of various major events which occur during the life of the logon session. For example, through the Telco or ISP the user is capable of connecting with various services (e.g. the Internet, private intra nets, private pay-for-access domains). Additionally, once the user has initialized or connected to the service, the individual PPP connections and PPP disconnections in to and out of the service can be accounted for.

Such multiple-level accounting would provide the Telco or ISP with flexibility in devising sophisticated rate schemes. The ISPs and Telcos would no longer be restricted by rate schemes solely based on account logon and account logoff, but rather customers could be charged in accordance to the specific services which they access and the duration, byte-count or quantity of the connections to those service. The ability to account for service establishment and PPP connections allows Telcos and ISPs to offer their customers (i.e. computer users) cost effective access to communities of interest (i.e. those domains sites which are designated as pay-per-use.) Telcos and ISPs would be afforded the capability to provide detailed billing information and create various service options.

BRIEF DESCRIPTION OF THE INVENTION

The present invention is a method and apparatus for providing computer network access points the capability for multiple-level accounting. A gateway device located at the access point is capable of generating Internet protocol accounting start and stop requests based on various events that are to be accounted for when a user accesses a network. These events include the user account logon, the service establishments and the Point to Point protocol (PPP) connections between the gateway device and public and private domains within the network. The counter is capable of tracking the duration of sessions and connections and the byte-count associated with the specified session or connection. The gateway device communicates with an accounting server which stores the accounting requests and matches start requests with subsequent stop requests.

OBJECTS AND ADVANTAGES OF THE INVENTION

Accordingly, it is an object and advantage of the present invention to provide a method for network access point maintainers to implement multiple-level accounting which thereby provides the capability for detailed billing records and create various service options.

Another object and advantage of the present invention is to provide for a accounting system within a PoP of a computer network which is capable of multiple-level accounting.

Another object and advantage of the present invention is to provide a computer network service provider the capability to account for account logons, service establishments and PPP connections.

Another object and advantage of the present invention is to provide a secured means for providing multiple-level accounting by implementing means for retrying accounting requests and providing for secondary hardware back-up.

Yet another object and advantage of the present invention is to allow for the accounting request packets to use eight byte integers thus operating at 1.5 meg bits/second and eliminating the concern over counter overflow.

These and many other objects and advantages of the present invention will become apparent to those of ordinary skill in the art from a consideration of the drawings and ensuing description.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a schematic drawing of a computer network embodying a method for multiple-level Internet protocol accounting in accordance with a presently preferred embodiment of the present invention.

FIGS. 2A and 2B are a flow diagram of a method for multiple-level Internet protocol accounting in a computer network in accordance with a presently preferred embodiment of the present invention.

FIG. 3 is a schematic of an accounting system within a computer network which employs multiple-level Internet protocol accounting in accordance with a presently preferred embodiment of the present invention.

FIG. 4 is a schematic of a computer network system which employs multiple-level Internet protocol accounting in accordance with a presently preferred embodiment of the present invention.

DETAILED DESCRIPTION OF THE PRESENT INVENTION

Those of ordinary skill in the art will realize that the following description of the present invention is illustrative only and is not intended to be in any way limiting. Other embodiments of the invention will readily suggest themselves to such skilled persons from an examination of the within disclosure.

The present invention allows the computer network service provider to assemble accounting records for three separate accounting events. These events are the initial network account logon/logoff, the service establishment/termination and the individual connection starts and stops to a specified service. The present invention is capable of accounting for both the time duration for an event and for the byte count encountered during a given event.

In a presently preferred embodiment of the present invention, a method for multiple-level internet protocol accounting is illustrated by the computer network schematic of FIG. **1** and the flow diagrams of FIGS. **2A** and **2B**. In the simplified schematic of a computer network **10** shown in FIG. **1**, the host **12** is able to connect with various private and public network domains **14** and **16**, including the internet through an access point **18**. The access point **18** is typically a computer network service provider, such as a telephone company (Telco) or commercial internet service provider (ISP). The access point serves as a link in the overall network scheme and houses various network interfaces and service components capable of routing and transferring data to and from various points on the network. Shown in FIG. **1** are a service selection gateway (SSG) **20**, such as the Cisco model 6510, manufactured by Cisco Systems, Inc. of San Jose, Calif. and an authentication, authorization and accounting (AAA) server **22**, such as Cisco ACS or Cisco Secure, manufactured by Cisco Systems, Inc. of San Jose, Calif. These devices are located within the access point **18** and are instrumental in carrying out the multiple-level internet protocol accounting method of this presently preferred embodiment. The AAA server **22** may accommodate several client SSG's simultaneously and communicate with one another according to a standard Internet protocol. For the presently preferred embodiment of this invention, the Remote Authentication Dial-In User Service (RADIUS) protocol is used as the communication protocol between the SSG **20** and the AAA server **22**. Those of ordinary skill in the art will realize that other internet protocols can be used as acceptable communication means between the various communication devices which encompass the computer network **10**.

At step **30** of FIG. **2A**, the user of the host computer **12** initiates a dashboard application program as a means of gaining access to a desired computer network. The dashboard application program will typically require the user to enter some form of user identification and authentication information, most generally, a user-name and a private password. It may be possible for the host computer to store such information in memory and provide this information to the application program automatically upon initiating the program. The application program will then contact a computer network service provider, typically a telephone company (Telco) or commercial internet service provider (ISP), via a modem and telephone line. At step **32**, the host **12** sends to the computer network service provider a RADIUS logon request data packet containing the user identification and authentication information. This data packet can come directly from the host computer **12** or, it is also within the inventive concept herein disclosed, to have this packet sent from an external web server.

At step **34**, the computer network service provider receives the RADIUS logon request data packet at the SSG **20** and initiates a RADIUS account logon access request packet for host authentication. At step **36**, the RADIUS account logon access request packet is directed to the AAA server **22** where, at step **38**, the access request packet is matched against unique user profiles in memory to verify the authenticity of the user host **12**. Next, at step **40** the AAA server **22** sends an authentication reply back to the SSG **20** which confirms the authentication and lists the services available for a particular user. If the authentication reply indicates that the authentication process was successful then, at step **42**, the SSG **20** logons on the host **12** by sending an access-accept packet from the SSG **20** to the host **12** informing the user that the logon process has been completed and displaying the available services on the user's dashboard. If the authentication reply indicates that the authentication process was unsuccessful then, at step **44**, the SSG **20** sends a prompt back to the host **12** notifying the user that the authentication process was unsuccessful.

Once the SSG **20** logons on the host **12**, at step **46** the SSG **20** generates an account logon accounting start request and, at step **48**, this accounting start request is sent to the AAA server **22**. In the preferred embodiment the RADIUS account logon accounting start request will have the following attributes associated with the record:

Acct-Status-Type=Start
NAS-IP-Address=ip_address
User-Name="username"
Acct-Session-Id="session_id"
Framed-IP-Address=user_ip
Proxy-State="n"
where:
ip_address=IP address of the SSG interface card 1.
username=Name used to log on to the service provider network
session_id=Session Number
user_ip=IP address of the user's system
n=Accounting record queuing information

Once the account logon accounting start request has been completed, at step **50**, the host **12** is capable of initiating a service logon request. At step **52**, the service logon request is sent from the host **12** to the computer network service provider where it is received by the SSG **20**. The SSG **20** makes an initial determination, at step **54**, to determine if the requested service profile is available within the SSG **20**. If the service profile pre-exists in the SSG **20** then a determination is made, at step **64**, as to whether the user has authorization to access the desired service. If no service profile exists, then at step **56**, the SSG **20** prepares a RADIUS service access request packet as a means for

verifying the authorized services available for the given user. The RADIUS service access request packet is sent, at step **58**, to the AAA server **22** where the information in the service access request packet is used to locate, at step **60**, service profile stored in the memory of the AAA server **22**. Once the service profile is found, it is forwarded to the SSG **20** at step **62** and at step **64** a determination is made as to whether the user has authorization to access the desired service. If no match to the desired service is found within the user profile then authorization to access the service is withheld and, at step **66**, a prompt is sent to the host **12** informing the user that service authorization is denied. If the service to which the user **12** desires access to is found within the user profile then the SSG **20** at step **68**, logs the user on to the desired service. The connection to the service may be accomplished by various means, including but not limited to, an L2TP (Layer Two Tunneling Protocol) tunnel connection **24** or a standard Internet (packet-forward) connection **26** via a leased line.

Once the service logon is successfully completed, at step **70**, the SSG **20** determines whether a pre-existing service connection exists (typically, generated by any other host which connects through this particular access point). If a service connection is pre-existing then no need exists to generate a service accounting start request packet because the AAA server **22** will already have such stored in memory. If no pre-existing service connection exists, then at step **72**, the SSG **20** generates a service accounting start request packet and at step **74** this request is sent from the SSG **20** to the AAA server **22**. In the preferred embodiment the RADIUS service accounting start request packet will have the following attributes associated with the record:

Acct-Status-Type=Start
NAS-IP-Address=ip_address
User-Name="service"
Acct-Session-Id="session_id"
Proxy-State="n"
where:
service=Name of the service profile.
ip_address=IP address of the SSG interface card 1.
session_id=Session Number.
n=Accounting record queuing information.

Once the service establishment is completed, the user, at step **76**, is connected with the home gateway **28** of the desired service. When this type of Point to Point Protocol (PPP) connection is made with the home gateway **78** of the desired service, it triggers the SSG **20**, at step **80**, to generate a connection accounting request packet and forward this request to the AAA server **22**, step **80**. In the preferred embodiment the RADIUS connection accounting start request packet will have the following attributes associated with the record:

Acct-Status-Type=Start
NAS-IP-Address=ip_address
ser-Name="username"
Acct-Authentic=RADIUS
Acct-Session-Id="session_id"
Service-Info="service"
Service-Info="hg_username"
Service-Info="type"
Proxy-State="n"
where:
ip_address=IP address of the SSG interface card 1.
username=Name used to log on to the service provider network
session_id=Session Number
service=Name of the service profile.
hg_username=The username used to authenticate the user with the home gateway.
type=TT—Tunneled connection.
TI—Internet (packet-forward) connection.
n=Accounting record queuing information.

Once a user desires to disconnect from an open PPP connection and, at step **82**, the disconnect is executed, the SSG **20** generates a connection accounting stop request packet, step **84**. While these packets will typically be generated due to a host-request disconnect command, it is also possible to generate the packets when a disconnect is prompted by other events outside the control of the user, such as, a lost-carrier, a lost-service or a session-timeout. Once the connection account stop request packet is generated, at step **86**, the packet is forwarded to the AAA server **22** where it is coupled with the initial connection start request for accounting and filing purposes. In the preferred embodiment the RADIUS connection accounting stop request packet will have the following attributes associated with the record:

Acct-Status-Type=Stop
NAS-IP-Address=ip_address
User-Name="username"
Acct-Input-Octets=in-bytes
Acct-Output-Octets=out_bytes
Acct-Session-Time=time
Acct-Terminate-Cause=cause
Acct-Session-Id="session_id"
Service-Info="service"
Service-Info="hg_username"
Service-Info="type"
Proxy-State="n"
where:
ip_address=IP address of the SSG interface card 1.
username=Name used to log on to the service provider network.
in_bytes=Number of inbound bytes.
out_bytes=Number of outbound bytes.
time=Length of session in seconds.
cause=Cause of account termination. These include:
—Lost-Carrier—Lost-Service—User-Request—Session-Timeout
session_id=Session Number.
service=Name of the service profile.
hg_username=The username used to authenticate the user with the home gateway.
type=TT—Tunneled connection.
TI—Internet (packet-forward) connection.
n=Accounting record queuing information.

Once the host terminates the PPP connection, at step **88**, the SSG **20** determines whether any remaining connections exist to the service. If the host **12** or other hosts connected through this access point have current connections existing with this service, then no need exists at this time to generate a service accounting stop request packet. The service accounting stop request packet will be generated only when no existing connections exist through the service connection. If the SSG **20** determines that no other then-existing connections exist through the service connection then, at step **90**, the SSG **20** generates a service accounting stop request packet. While these service stop request packets will typically be generated as a result of a host-request disconnect command, it is also possible to generate the packets

when a disconnect is prompted by other events outside the control of the user, such as, a lost-carrier, a lost-service or a session-timeout. Once the service account stop request packet is generated, at step **92**, the packet is forwarded to the AAA server **22** where it is coupled with the initial service start request for accounting and filing purposes. In the preferred embodiment the RADIUS service accounting stop request packet will have the following attributes associated with the record:

Acct-Status-Type=Stop
NAS-IP-Address=ip_address
User-Name="service"
Acct-Input-Octets=in-bytes
Acct-Output-Octets=out_bytes
Acct-Session-Time=time
Acct-Terminate-Cause=cause
Acct-Session-Id="session_id"
Proxy-State="n"
where:
service=Name of the service profile.
ip_address=IP address of the SSG interface card 1.
in_bytes=Number of inbound bytes.
out_bytes=Number of outbound bytes.
time=Length of session in seconds.
cause=Cause of account termination. These include:
—Lost-Carrier—Lost-Service—Host-Request—Session-Timeout
session_id=Session Number.
n=Accounting record queuing information.

Finally, once the host desires to logoff from the network account and, at step **94**, the logoff is executed, the SSG **20** generates an account logoff accounting stop request packet, step **96**. While these packets will typically be generated due to a user-requested logoff command, it is also possible to generate the stop packets when a logoff is prompted by other events outside the control of the user, such as a session timeout. Once the account logoff accounting stop request packet is generated, at step **98**, the packet is forwarded to the AAA server **22** where it is coupled with the initial account logon accounting start request for further accounting and filing purposes. In the preferred embodiment the RADIUS account logoff accounting stop request packet will have the following attributes associated with the record:

Acct-Status-Type=Stop
NAS-IP-Address=ip_address
User-Name="username"
Acct-Session-Time=time
Acct-Terminate-Cause=cause
Acct-Session-Id=user_id
framed-IP-Address=user_ip
Proxy-State="n"
where:
ip_address=IP address of the SSG interface card 1.
username=Name used to log on to the service provider network.
time=Length of session in seconds.
cause=Cause of account termination. These include:
—User-Request—Session-Timeout
session_id=Session Number.
user_ip=IP address of the user's system.
n=Accounting record queuing information.

In another preferred embodiment of the present invention the method for multiple-level accounting can include an additional accounting session implemented to provide periodic update information to the AAA server **22**. These periodic accounting update request packets contain information which, typically, mirrors the data found in account logoff accounting stop request packets and serve as an added measure of security should system errors, transmission errors, or the like, prevent the actual account logoff accounting stop request from being either properly generated at the SSG **20** or properly sent to the AAA server **22**. The Telco or commercial ISP is capable of defining the attributes associated with the periodic accounting update request and defining the time period between such requests.

Additional security measures can also be imposed by the SSG **20** to insure that accounting request packets are properly received by the AAA server **22**. These security measures include defining within the SSG **20** scheme a retry value and an interval between retry to be employed when the SSG **20** makes a determination that an accounting request packet was not properly delivered to the AAA **22** server. The retry value and the interval between retry are SSG-user configurable.

In another embodiment of the present invention, a networking event metering system using the multiple level accounting of the present invention is detailed in FIG. **3**. In this configuration the networking event metering system **120** includes a gateway device **122**. The gateway device **122** is in communication with a plurality of hosts **124**, typically by means of a telephone line. Those of ordinary skill in the art will recognize that other types of host-to-gateway device access methods may be provided by a Telcos or ISP such as frame relay, leased lines, ATM (Asynchronous Transfer Mode), ADSL (Asymmetric Digital Subscriber Line) and the like. The gateway device **122** has the capability to process access requests being sent from the hosts. These access requests would include, but not be limited to, account logon requests, service authorization requests and connection requests. In a preferred embodiment, the gateway device implements the RADIUS protocol as the communication language between itself and other ISP network interfaces. The gateway device **122** receives the account logon access requests from the hosts **124** and the proxier **126** within the gateway device **122** correspondingly proxies these access requests to AAA server **128** for user authentication purposes.

The AAA server **128** then compares attribute data found in the account logon access request packets against data found in the user profiles **130** of the AAA server's memory bank. If the data in the RADIUS access requests are found to match data in the AAA server **128** then access to the network can be granted. Additionally, the AAA server **128** holds service profiles **132** within the memory bank. The user profiles **130** contain, among various data attributes, a listing of which services (private and public domains) a specified user is authorized to access. The service profiles **132**, which are not user dependant, contain attribute date for a specific service domain. The gateway device **122** can query the AAA server **128** for a given user profile **130** and then within the processor **134** of the gateway device **122** assess the profile to determine which services the user has authorization to access.

The gateway device **122** also encompasses an accounting request generator **136** which is capable of generating the accounting start and stop requests for various events which are triggered within the gateway device **122**. Upon the gateway device **122** authenticating the user for account logon, the accounting request generator **136** will issue an account logon accounting start request packet and forward the packet to the accounter **138** located within the AAA server **128**. Similarly, upon the gateway device **122** autho-

rizing the user to access a specific service, the accounting request generator **136** will, typically, issue a service accounting start request and forward the packet to the accounter **138** located within the AAA server **128**. In the instance where the service already has a then-existing connection established by the host or any other host connecting through the gateway device **122**, no need would exist to generate the service start request because such a request will already be existing within the accounter **138**. Once the user establishes a PPP connection **140** to the desired service **142**, the accounting request generator **136** will issue a PPP connection accounting start request and forward the packet to the accounter **138** located within the AAA server **128**.

The gateway device **122** will also rely on the accounting request generator **136** to issue the accounting stop requests upon certain triggering events occurring. When the user disconnects from a PPP connection, the accounting request generator **136** will issue a PPP connection accounting stop request packet and forward this packet to the accounter **138** where it will be married with its corresponding start request for accounting purposes. While these packets will typically be generated due to a host-request disconnect command, it is also possible to generate the packets when a disconnect is prompted by other events outside the control of the user, such as, a lost-carrier, a lost-service or a session-timeout. When the user disconnects from a PPP connection the processor **134** within the gateway device **122** will determine if any additional connections remain to that particular service through that particular tunnel or routed connection. If no additional connections remain, then the accounting request generator **136** will issue a service accounting stop request packet and forward this packet to the accounter **138** where it will be married with the corresponding stop request for subsequent accounting purposes. While these service stop request packets will typically be generated as a result of a host-request disconnect command, it is also possible to generate the packets when a disconnect is prompted by other events outside the control of the user, such as, a lost-carrier, a lost-service or a session-timeout. Finally, when the host initiates an account logoff, the accounting request generator **136** will issue an account logoff accounting stop request packet and forward this to the accounter **138** where it will be married with the corresponding stop request for subsequent accounting purposes. While these packets will typically be generated due to a user-requested logoff command, it is also possible to generate the stop packets when a logoff is prompted by other events outside the control of the user, such as a session timeout.

Additionally, it is within the inventive concept herein disclosed to provide back-up capacity to both the gateway device **122** and AAA server **128** to compensate for device failures or errors in transmission. A back-up gateway device **144** allows for a safeguard should the primary gateway device **122** become temporarily inactive and fail to provide a means for generating and sending accounting start or stop requests. a secondary AAA server **146** allows for the initial accounting start request data to be catalogued in a second location should the primary AAA server encounter a temporary failure or service outage.

In another preferred embodiment of the present invention, a computer network having the capabilities to use multiple-level accounting is illustrated in FIG. **4**. A plurality of hosts **150** have the capacity to gain network access through network access points **152**. The access point is typically operated by a Telco or ISP and houses various network interfaces and service components. Included among these components are a plurality of gateway devices **154** and AAA servers **156**. The AAA servers **156** may accommodate several client SSG's simultaneously and communicate with one another according to a standard Internet protocol, such as RADIUS. The gateway devices **154** are in communication with a plurality of hosts **150**, typically by means of a telephone line. Those of ordinary skill in the art will recognize that other types of host-to-gateway device access methods may be provided by a Telcos or ISP such as frame relay, leased lines, ATM (Asynchronous Transfer Mode), ADSL (Asymmetric Digital Subscriber Line) and the like. The gateway devices **154** are capable of receiving account logon and service requests from the hosts **150**, proxying these requests to the AAA servers **156** for authentication and authorization and then determining account logon and service. Once these determinations are made the gateway devices generates accounting start requests based on account logon, service establishment and PPP connection start. When the host initiates a termination request or when other events outside the control of the user dictate such, the gateway device generates accounting stop requests based on account logoff, service disconnect and PPP connection stop.

Alternative Embodiments

Although illustrative presently preferred embodiments and applications of this invention are shown and described herein, many variations and modifications are possible which remain within the concept, scope and spirit of the invention, and these variations would become clear to those skilled in the art after perusal of this application. The invention, therefore, is not limited except in spirit of the appended claims.

What is claimed is:

1. A method for providing multiple-level accounting to a computer network service provider comprising the steps of:

generating an internet protocol account logon accounting start request in response to a determination that a subscriber has logged on to said computer network service provider;

forwarding said internet protocol account logon start accounting request to a memory;

generating an internet protocol service accounting start request in response to a determination that said subscriber has logged on to a service and that no other subscribers have a current connection established to said service through the computer network service provider interface;

forwarding said internet protocol service accounting start request to a memory;

generating an internet protocol service accounting stop request in response to a determination that said subscriber has terminated said service and that no other subscribers have a current connection established to said service through the computer network service provider interface; and

forwarding said internet protocol service accounting stop request to a memory.

2. The method of claim **1** further comprising the steps of:

generating an internet protocol connection accounting start request in response to a determination that said subscriber has established a connection with said service;

forwarding said internet protocol connection accounting start requests to a memory;

generating an internet protocol connection accounting stop request in response to a determination that said subscriber has terminated said connection with said service; and

forwarding said internet protocol connection accounting stop requests to a memory.

3. The method of claim 1 further comprising the steps of:

generating an internet protocol account logoff accounting stop request in response to a determination that the subscriber has logged off said computer network service provider; and

forwarding said internet protocol account logoff accounting stop requests to a memory.

4. The method of claim 1 further comprising the steps of:

generating an internet protocol update accounting request at predetermined intervals after said internet protocol account logon accounting start request has been received into said memory; and

forwarding said internet protocol update accounting requests to a memory.

5. A method for providing multiple-level accounting to a computer network service provider comprising the steps of:

receiving a user account logon access request at said computer network service provider;

authorizing and authenticating said user in response to receiving said user account logon request;

sending an account logon accounting start request to an accounting server;

receiving a user service access request at said computer network service provider to access a requested service;

granting user service privileges to said user if said requested service is contained within a service profile of said user;

sending a service accounting start request to said accounting server if said computer network service provider interface has not already established a service connection with said requested service; and

responsive to a determination that said user has disconnected from said requested service, sending a service accounting stop request to said accounting server if no other hosts have current connections to said requested service.

6. The method of claim 5 further comprising the steps of:

establishing a connection between said user and said requested service after granting user service privileges to said user;

sending a connection accounting start request to said accounting server; and

responsive to a determination that said user has disconnected from said requested service, sending a connection accounting stop request to said accounting server.

7. The method of claim 6 further comprising the steps of:

responsive to a determination that said user has terminated said account logon, sending an account logoff accounting stop request to said accounting server.

8. The method of claim 7 further comprising the step of:

re-sending said account logoff accounting stop request at predetermined intervals upon determining that said initial connection accounting stop request was not received by said accounting server.

9. The method of claim 6 further comprising the steps of:

re-sending said connection accounting start request at predetermined intervals upon determining that said initial connection accounting start request was not received by said accounting server; and

re-sending said connection accounting stop request at predetermined intervals upon determining that said initial connection accounting stop request was not received by said accounting server.

10. The method of claim 5 wherein the step of receiving a user account logon access request further comprises receiving said request directly from a dashboard application launched on said user's host computer.

11. The method of claim 5 wherein the step of receiving a user account logon access request further comprises receiving said request from a remote web server.

12. The method of claim 5 further comprising the steps of:

generating update accounting requests at predetermined intervals after said account logon accounting start request has been received into said accounting server; and

forwarding said internet protocol update accounting request to said accounting server.

13. The method of claim 5 further comprising the steps of:

re-sending said account logon accounting start request at predetermined intervals upon determining that said initial account logon accounting start request was not received by said accounting server;

re-sending said service accounting start request at predetermined intervals upon determining that said initial service accounting start request was not received by said accounting server; and

re-sending said service accounting stop request at predetermined intervals upon determining that said initial service accounting stop request was not received by said accounting server.

14. An accounting metering apparatus for providing multiple-level accounting to a computer network service provider comprising the following:

a means for receiving a user account logon access request at said computer network service provider;

a means for authorizing and authenticating said user in response to receiving said user account logon request;

a means for sending an account logon accounting start request to an accounting server;

a means for receiving a user service access request at said computer network service provider to access a requested service;

a means for granting user service privileges to said user if said requested service is contained within a service profile of said user;

a means for sending a service accounting start request to said accounting server if said computer network service provider interface has not already established a service connection with said requested service; and

a means for, responsive to a determination that said user has disconnected from said requested service, sending a service accounting stop request to said accounting server if no other hosts have current connections to said requested service.

15. The accounting metering apparatus of claim 14 further comprising the following:

a means for establishing a connection between said user and said requested service after granting user service privileges to said user;

a means for sending a connection accounting start request to said accounting server; and

a means for, responsive to a determination that said user has disconnected from said requested service, sending a connection accounting stop request to said accounting server.

16. The accounting metering apparatus of claim 15 further comprising the following:

a means for, responsive to a determination that said user has terminated said account logon, sending an account logoff accounting stop request to said accounting server.

17. An accounting metering system for providing multiple-level accounting to a network service provider within a computer network comprising the following:

a gateway device capable of receiving network, service and connection access request packets from a plurality of hosts, said gateway device including:

a proxier capable of readying logon authentication request packets and service authorization request packets for verification,

a processor capable of granting network access, service establishment and connection access;

an accounting request generator capable of generating accounting start request packets and accounting stop request packets for network account logons and logoffs, service establishments and terminations and connection accesses and terminations;

an authentication, authorization and accounting server communicating with said gateway device by a network service provider proscribed internet protocol, said authentication, authorization and accounting server consisting of:

an authenticator capable of verifying the authenticity of a user based on information contained in memory and comparable information found in said logon authentication request packets;

a user specific service profile which contains a listing of the services which the user is authorized to access; and

an accounting tabulator capable of storing said accounting start requests and said accounting stop requests received from said gateway device.

18. The accounting metering system of claim 17 wherein said internet protocol used as a communication means between said gateway device and said authentication, authorization and accounting server further comprises the RADIUS protocol.

19. The accounting metering system of claim 17 wherein said accounting start requests and said accounting stop requests further comprise account attributes for both time and byte counts.

20. The accounting system of claim 17 further comprising:

a second gateway device capable of acting as a standby gateway device if the first gateway device fails or becomes temporarily non-functional.

21. The accounting system of claim 17 further comprising:

a second authentication, authorization and accounting device capable of acting as a standby authentication, authorization and accounting device if the first authentication, authorization and accounting device fails or becomes temporarily non-functional.

22. The accounting system of claim 17 wherein said network, service and connection access request packets and said network, service and connection accounting request packets further comprise eight byte integer counters.

* * * * *

Network Working Group
Request for Comments: 2866
Category: Informational
Obsoletes: 2139

C. Rigney
Livingston
June 2000

RADIUS Accounting

Status of this Memo

This memo provides information for the Internet community. It does not specify an Internet standard of any kind. Distribution of this memo is unlimited.

Copyright Notice

Abstract

This document describes a protocol for carrying accounting information between a Network Access Server and a shared Accounting Server.

Implementation Note

This memo documents the RADIUS Accounting protocol. The early deployment of RADIUS Accounting was done using UDP port number 1646, which conflicts with the "sa-msg-port" service. The officially assigned port number for RADIUS Accounting is 1813.

Table of Contents

1. Introduction

Managing dispersed serial line and modem pools for large numbers of users can create the need for significant administrative support. Since modem pools are by definition a link to the outside world, they require careful attention to security, authorization and accounting. This can be best achieved by managing a single "database" of users, which allows for authentication (verifying user name and password) as well as configuration information detailing the type of service to deliver to the user (for example, SLIP, PPP, telnet, rlogin).

The RADIUS (Remote Authentication Dial In User Service) document [2] specifies the RADIUS protocol used for Authentication and Authorization. This memo extends the use of the RADIUS protocol to cover delivery of accounting information from the Network Access Server (NAS) to a RADIUS accounting server.

This document obsoletes RFC 2139 [1]. A summary of the changes between this document and RFC 2139 is available in the "Change Log" appendix.

Key features of RADIUS Accounting are:

Client/Server Model

A Network Access Server (NAS) operates as a client of the RADIUS accounting server. The client is responsible for passing user accounting information to a designated RADIUS accounting server.

The RADIUS accounting server is responsible for receiving the accounting request and returning a response to the client indicating that it has successfully received the request.

The RADIUS accounting server can act as a proxy client to other kinds of accounting servers.

Network Security

Transactions between the client and RADIUS accounting server are authenticated through the use of a shared secret, which is never sent over the network.

Extensible Protocol

All transactions are comprised of variable length Attribute-Length-Value 3-tuples. New attribute values can be added without disturbing existing implementations of the protocol.

1.1. Specification of Requirements

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [3]. These key words mean the same thing whether capitalized or not.

1.2. Terminology

This document uses the following terms:

service The NAS provides a service to the dial-in user, such as PPP or Telnet.

session Each service provided by the NAS to a dial-in user constitutes a session, with the beginning of the session defined as the point where service is first provided and the end of the session defined as the point where service is ended. A user may have multiple sessions in parallel or series if the NAS supports that, with each session generating a separate start and stop accounting record with its own Acct-Session-Id.

silently discard
This means the implementation discards the packet without further processing. The implementation SHOULD provide the capability of logging the error, including the contents of the silently discarded packet, and SHOULD record the event in a statistics counter.

2. Operation

When a client is configured to use RADIUS Accounting, at the start of service delivery it will generate an Accounting Start packet describing the type of service being delivered and the user it is being delivered to, and will send that to the RADIUS Accounting server, which will send back an acknowledgement that the packet has been received. At the end of service delivery the client will generate an Accounting Stop packet describing the type of service that was delivered and optionally statistics such as elapsed time, input and output octets, or input and output packets. It will send that to the RADIUS Accounting server, which will send back an acknowledgement that the packet has been received.

The Accounting-Request (whether for Start or Stop) is submitted to the RADIUS accounting server via the network. It is recommended that the client continue attempting to send the Accounting-Request packet until it receives an acknowledgement, using some form of backoff. If no response is returned within a length of time, the request is re-sent a number of times. The client can also forward requests to an alternate server or servers in the event that the primary server is down or unreachable. An alternate server can be used either after a number of tries to the primary server fail, or in a round-robin fashion. Retry and fallback algorithms are the topic of current research and are not specified in detail in this document.

The RADIUS accounting server MAY make requests of other servers in order to satisfy the request, in which case it acts as a client.

If the RADIUS accounting server is unable to successfully record the accounting packet it MUST NOT send an Accounting-Response acknowledgment to the client.

2.1. Proxy

See the "RADIUS" RFC [2] for information on Proxy RADIUS. Proxy Accounting RADIUS works the same way, as illustrated by the following example.

1. The NAS sends an accounting-request to the forwarding server.

2. The forwarding server logs the accounting-request (if desired), adds its Proxy-State (if desired) after any other Proxy-State attributes, updates the Request Authenticator, and forwards the request to the remote server.

3. The remote server logs the accounting-request (if desired), copies all Proxy-State attributes in order and unmodified from the request to the response packet, and sends the accounting-response to the forwarding server.

4. The forwarding server strips the last Proxy-State (if it added one in step 2), updates the Response Authenticator and sends the accounting-response to the NAS.

A forwarding server MUST not modify existing Proxy-State or Class attributes present in the packet.

A forwarding server may either perform its forwarding function in a pass through manner, where it sends retransmissions on as soon as it gets them, or it may take responsibility for retransmissions, for example in cases where the network link between forwarding and remote server has very different characteristics than the link between NAS and forwarding server.

Extreme care should be used when implementing a proxy server that takes responsibility for retransmissions so that its retransmission policy is robust and scalable.

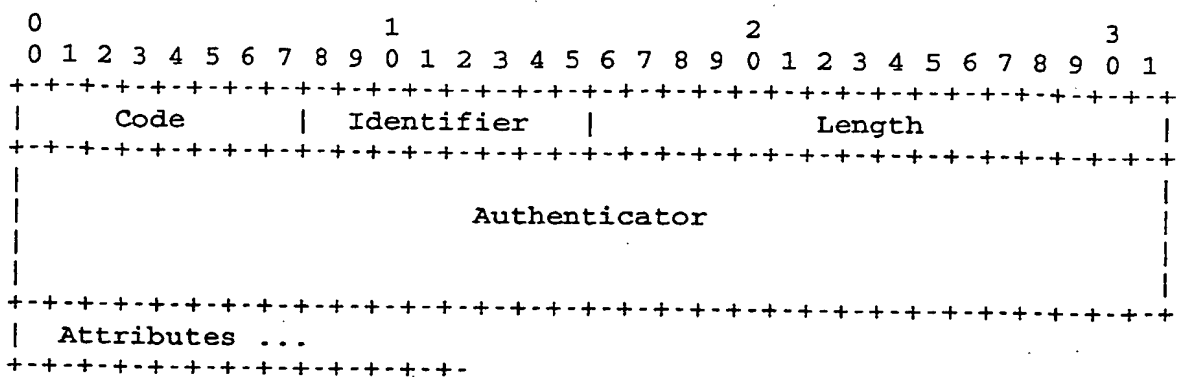
3. Packet Format

Exactly one RADIUS Accounting packet is encapsulated in the UDP Data field [4], where the UDP Destination Port field indicates 1813 (decimal).

When a reply is generated, the source and destination ports are reversed.

This memo documents the RADIUS Accounting protocol. The early deployment of RADIUS Accounting was done using UDP port number 1646, which conflicts with the "sa-msg-port" service. The officially assigned port number for RADIUS Accounting is 1813.

A summary of the RADIUS data format is shown below. The fields are transmitted from left to right.



Code

The Code field is one octet, and identifies the type of RADIUS packet. When a packet is received with an invalid Code field, it is silently discarded.

RADIUS Accounting Codes (decimal) are assigned as follows:

4	Accounting-Request
5	Accounting-Response

Identifier

The Identifier field is one octet, and aids in matching requests and replies. The RADIUS server can detect a duplicate request if it has the same client source IP address and source UDP port and Identifier within a short span of time.

Length

The Length field is two octets. It indicates the length of the packet including the Code, Identifier, Length, Authenticator and Attribute fields. Octets outside the range of the Length field MUST be treated as padding and ignored on reception. If the packet is shorter than the Length field indicates, it MUST be silently discarded. The minimum length is 20 and maximum length is 4095.

Authenticator

The Authenticator field is sixteen (16) octets. The most significant octet is transmitted first. This value is used to authenticate the messages between the client and RADIUS accounting server.

Request Authenticator

In Accounting-Request Packets, the Authenticator value is a 16 octet MD5 [5] checksum, called the Request Authenticator.

The NAS and RADIUS accounting server share a secret. The Request Authenticator field in Accounting-Request packets contains a one-way MD5 hash calculated over a stream of octets consisting of the Code + Identifier + Length + 16 zero octets + request attributes + shared secret (where + indicates concatenation). The 16 octet MD5 hash value is stored in the Authenticator field of the Accounting-Request packet.

Note that the Request Authenticator of an Accounting-Request can not be done the same way as the Request Authenticator of a RADIUS Access-Request, because there is no User-Password attribute in an Accounting-Request.

Response Authenticator

The Authenticator field in an Accounting-Response packet is called the Response Authenticator, and contains a one-way MD5 hash calculated over a stream of octets consisting of the Accounting-Response Code, Identifier, Length, the Request Authenticator field from the Accounting-Request packet being replied to, and the response attributes if any, followed by the shared secret. The resulting 16 octet MD5 hash value is stored in the Authenticator field of the Accounting-Response packet.

Attributes

Attributes may have multiple instances, in such a case the order of attributes of the same type SHOULD be preserved. The order of attributes of different types is not required to be preserved.

4. Packet Types

The RADIUS packet type is determined by the Code field in the first octet of the packet.

4.1. Accounting-Request

Description

Accounting-Request packets are sent from a client (typically a Network Access Server or its proxy) to a RADIUS accounting server, and convey information used to provide accounting for a service provided to a user. The client transmits a RADIUS packet with the Code field set to 4 (Accounting-Request).

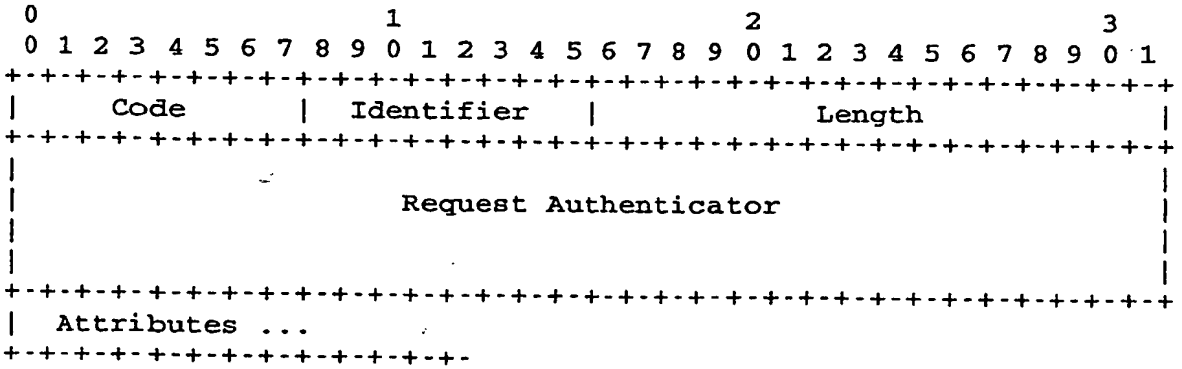
Upon receipt of an Accounting-Request, the server MUST transmit an Accounting-Response reply if it successfully records the accounting packet, and MUST NOT transmit any reply if it fails to record the accounting packet.

Any attribute valid in a RADIUS Access-Request or Access-Accept packet is valid in a RADIUS Accounting-Request packet, except that the following attributes MUST NOT be present in an Accounting-Request: User-Password, CHAP-Password, Reply-Message, State. Either NAS-IP-Address or NAS-Identifier MUST be present in a RADIUS Accounting-Request. It SHOULD contain a NAS-Port or NAS-Port-Type attribute or both unless the service does not involve a port or the NAS does not distinguish among its ports.

If the Accounting-Request packet includes a Framed-IP-Address, that attribute MUST contain the IP address of the user. If the Access-Accept used the special values for Framed-IP-Address telling the NAS to assign or negotiate an IP address for the user, the Framed-IP-Address (if any) in the Accounting-Request MUST contain the actual IP address assigned or negotiated.

A summary of the Accounting-Request packet format is shown below.

The fields are transmitted from left to right.



Code

4 for Accounting-Request.

Identifier

The Identifier field MUST be changed whenever the content of the Attributes field changes, and whenever a valid reply has been received for a previous request. For retransmissions where the contents are identical, the Identifier MUST remain unchanged.

Note that if Acct-Delay-Time is included in the attributes of an Accounting-Request then the Acct-Delay-Time value will be updated when the packet is retransmitted, changing the content of the Attributes field and requiring a new Identifier and Request Authenticator.

Request Authenticator

The Request Authenticator of an Accounting-Request contains a 16-octet MD5 hash value calculated according to the method described in "Request Authenticator" above.

Attributes

The Attributes field is variable in length, and contains a list of Attributes.

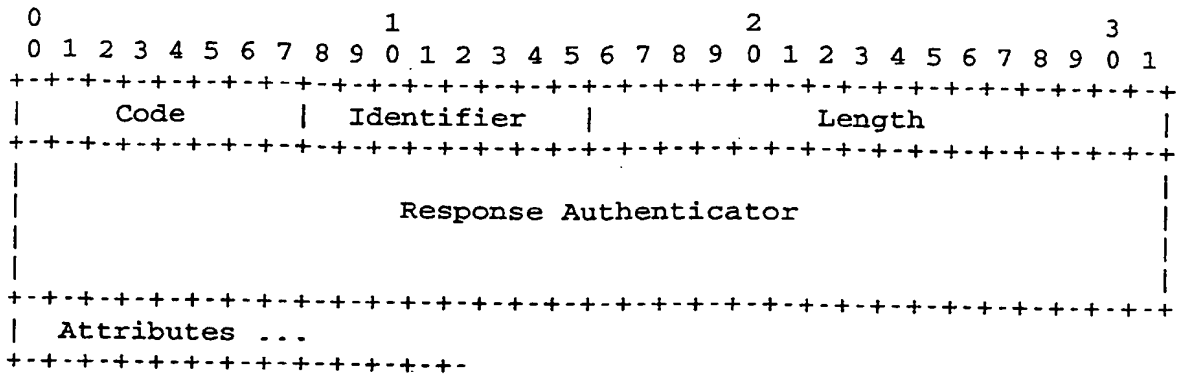
4.2. Accounting-Response

Description

Accounting-Response packets are sent by the RADIUS accounting server to the client to acknowledge that the Accounting-Request has been received and recorded successfully. If the Accounting-Request was recorded successfully then the RADIUS accounting server MUST transmit a packet with the Code field set to 5 (Accounting-Response). On reception of an Accounting-Response by the client, the Identifier field is matched with a pending Accounting-Request. The Response Authenticator field MUST contain the correct response for the pending Accounting-Request. Invalid packets are silently discarded.

A RADIUS Accounting-Response is not required to have any attributes in it.

A summary of the Accounting-Response packet format is shown below. The fields are transmitted from left to right.



Code

5 for Accounting-Response.

Identifier

The Identifier field is a copy of the Identifier field of the Accounting-Request which caused this Accounting-Response.

Response Authenticator

The Response Authenticator of an Accounting-Response contains a 16-octet MD5 hash value calculated according to the method described in "Response Authenticator" above.

Attributes

The Attributes field is variable in length, and contains a list of zero or more Attributes.

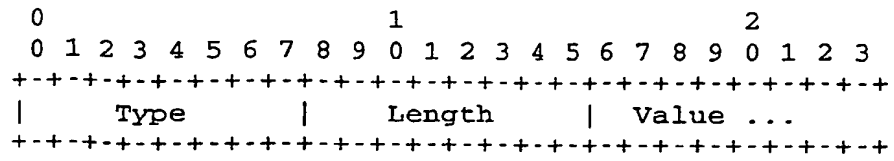
5. Attributes

RADIUS Attributes carry the specific authentication, authorization and accounting details for the request and response.

Some attributes MAY be included more than once. The effect of this is attribute specific, and is specified in each attribute description.

The end of the list of attributes is indicated by the Length of the RADIUS packet.

A summary of the attribute format is shown below. The fields are transmitted from left to right.



Type

The Type field is one octet. Up-to-date values of the RADIUS Type field are specified in the most recent "Assigned Numbers" RFC [6]. Values 192-223 are reserved for experimental use, values 224-240 are reserved for implementation-specific use, and values 241-255 are reserved and should not be used. This specification concerns the following values:

1-39	(refer to RADIUS document [2])
40	Acct-Status-Type
41	Acct-Delay-Time
42	Acct-Input-Octets
43	Acct-Output-Octets
44	Acct-Session-Id
45	Acct-Authentic
46	Acct-Session-Time
47	Acct-Input-Packets
48	Acct-Output-Packets
49	Acct-Terminate-Cause
50	Acct-Multi-Session-Id
51	Acct-Link-Count
60+	(refer to RADIUS document [2])

Length

The Length field is one octet, and indicates the length of this attribute including the Type, Length and Value fields. If an attribute is received in an Accounting-Request with an invalid Length, the entire request MUST be silently discarded.

Value

The Value field is zero or more octets and contains information specific to the attribute. The format and length of the Value field is determined by the Type and Length fields.

Note that none of the types in RADIUS terminate with a NUL (hex 00). In particular, types "text" and "string" in RADIUS do not terminate with a NUL (hex 00). The Attribute has a length field and does not use a terminator. Text contains UTF-8 encoded 10646

[7] characters and String contains 8-bit binary data. Servers and servers and clients MUST be able to deal with embedded nulls. RADIUS implementers using C are cautioned not to use strcpy() when handling strings.

The format of the value field is one of five data types. Note that type "text" is a subset of type "string."

text 1-253 octets containing UTF-8 encoded 10646 [7] characters. Text of length zero (0) MUST NOT be sent; omit the entire attribute instead.

string 1-253 octets containing binary data (values 0 through 255 decimal, inclusive). Strings of length zero (0) MUST NOT be sent; omit the entire attribute instead.

address 32 bit value, most significant octet first.

integer 32 bit unsigned value, most significant octet first.

time 32 bit unsigned value, most significant octet first -- seconds since 00:00:00 UTC, January 1, 1970. The standard Attributes do not use this data type but it is presented here for possible use in future attributes.

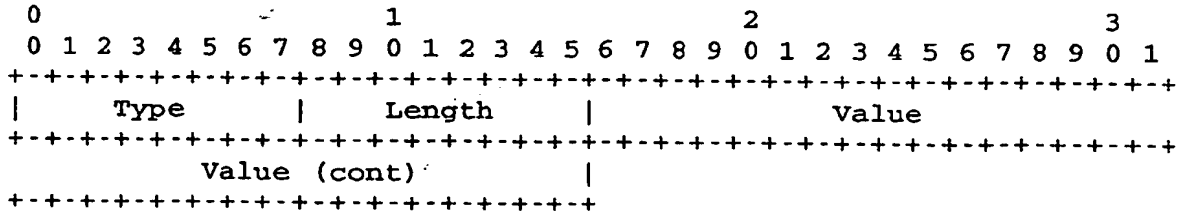
5.1. Acct-Status-Type

Description

This attribute indicates whether this Accounting-Request marks the beginning of the user service (Start) or the end (Stop).

It MAY be used by the client to mark the start of accounting (for example, upon booting) by specifying Accounting-On and to mark the end of accounting (for example, just before a scheduled reboot) by specifying Accounting-Off.

A summary of the Acct-Status-Type attribute format is shown below. The fields are transmitted from left to right.



Type

40 for Acct-Status-Type.

Length

6

Value

The Value field is four octets.

1	Start
2	Stop
3	Interim-Update
7	Accounting-On
8	Accounting-Off
9-14	Reserved for Tunnel Accounting
15	Reserved for Failed

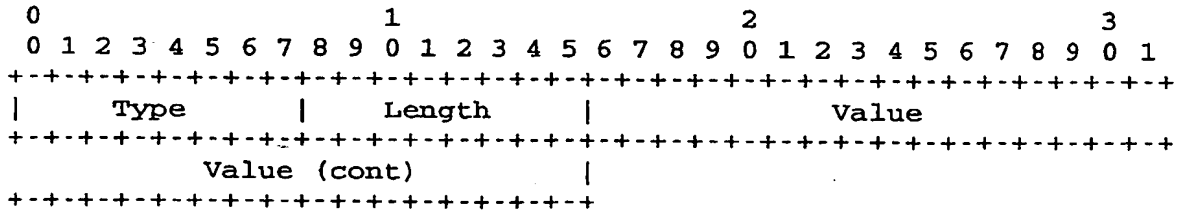
5.2. Acct-Delay-Time

Description

This attribute indicates how many seconds the client has been trying to send this record for, and can be subtracted from the time of arrival on the server to find the approximate time of the event generating this Accounting-Request. (Network transit time is ignored.)

Note that changing the Acct-Delay-Time causes the Identifier to change; see the discussion under Identifier above.

A summary of the Acct-Delay-Time attribute format is shown below. The fields are transmitted from left to right.



Type

41 for Acct-Delay-Time.

Length

6

Value

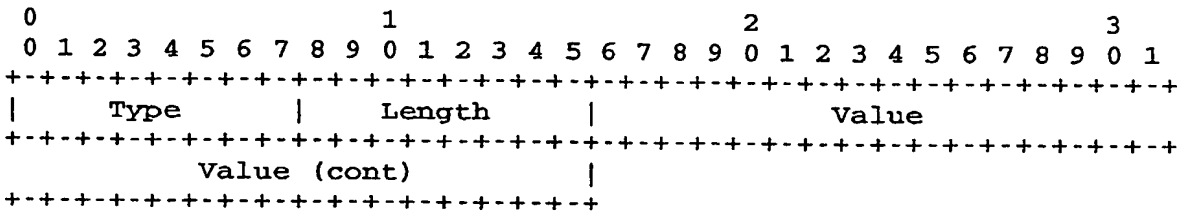
The Value field is four octets.

5.3. Acct-Input-Octets

Description

This attribute indicates how many octets have been received from the port over the course of this service being provided, and can only be present in Accounting-Request records where the Acct-Status-Type is set to Stop.

A summary of the Acct-Input-Octets attribute format is shown below. The fields are transmitted from left to right.



Type

42 for Acct-Input-Octets.

Length

6

Value

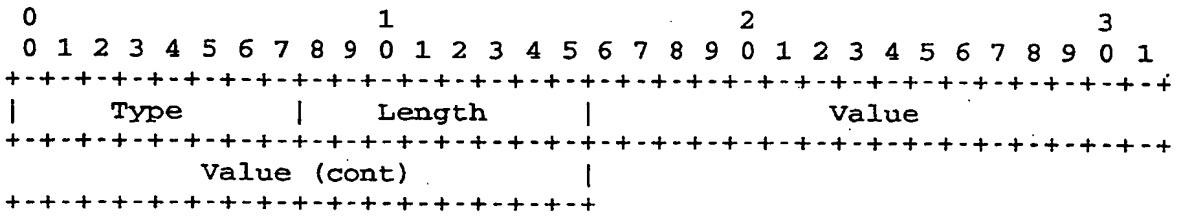
The Value field is four octets.

5.4. Acct-Output-Octets

Description

This attribute indicates how many octets have been sent to the port in the course of delivering this service, and can only be present in Accounting-Request records where the Acct-Status-Type is set to Stop.

A summary of the Acct-Output-Octets attribute format is shown below. The fields are transmitted from left to right.



Type

43 for Acct-Output-Octets.

Length

6

Value

The Value field is four octets.

5.5. Acct-Session-Id

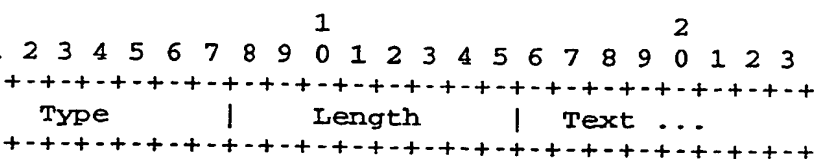
Description

This attribute is a unique Accounting ID to make it easy to match start and stop records in a log file. The start and stop records for a given session MUST have the same Acct-Session-Id. An Accounting-Request packet MUST have an Acct-Session-Id. An Access-Request packet MAY have an Acct-Session-Id; if it does, then the NAS MUST use the same Acct-Session-Id in the Accounting-Request packets for that session.

The Acct-Session-Id SHOULD contain UTF-8 encoded 10646 [7] characters.

For example, one implementation uses a string with an 8-digit upper case hexadecimal number, the first two digits increment on each reboot (wrapping every 256 reboots) and the next 6 digits counting from 0 for the first person logging in after a reboot up to 2^24-1, about 16 million. Other encodings are possible.

A summary of the Acct-Session-Id attribute format is shown below. The fields are transmitted from left to right.



Type

44 for Acct-Session-Id.

Length

>= 3

String

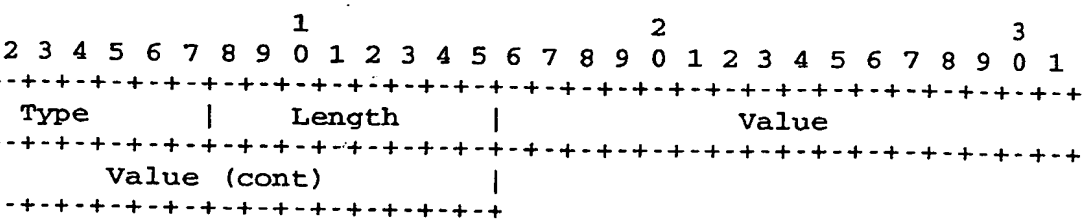
The String field SHOULD be a string of UTF-8 encoded 10646 [7] characters.

5.6. Acct-Authentic

Description

This attribute MAY be included in an Accounting-Request to indicate how the user was authenticated, whether by RADIUS, the NAS itself, or another remote authentication protocol. Users who are delivered service without being authenticated SHOULD NOT generate Accounting records.

A summary of the Acct-Authentic attribute format is shown below. The fields are transmitted from left to right.



Type

45 for Acct-Authentic.

Length

6

Value

The Value field is four octets.

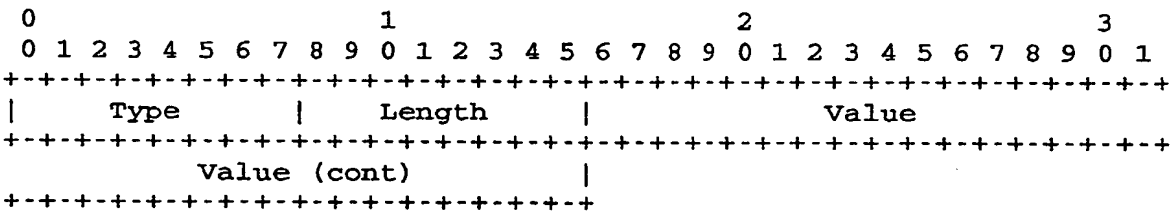
1	RADIUS
2	Local
3	Remote

5.7. Acct-Session-Time

Description

This attribute indicates how many seconds the user has received service for, and can only be present in Accounting-Request records where the Acct-Status-Type is set to Stop.

A summary of the Acct-Session-Time attribute format is shown below. The fields are transmitted from left to right.



Type

46 for Acct-Session-Time.

Length

6

Value

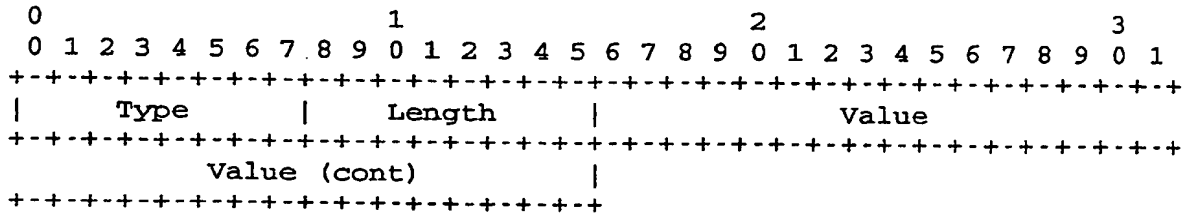
The Value field is four octets.

5.8. Acct-Input-Packets

Description

This attribute indicates how many packets have been received from the port over the course of this service being provided to a Framed User, and can only be present in Accounting-Request records where the Acct-Status-Type is set to Stop.

A summary of the Acct-Input-packets attribute format is shown below. The fields are transmitted from left to right.



Type

47 for Acct-Input-Packets.

Length

6

Value

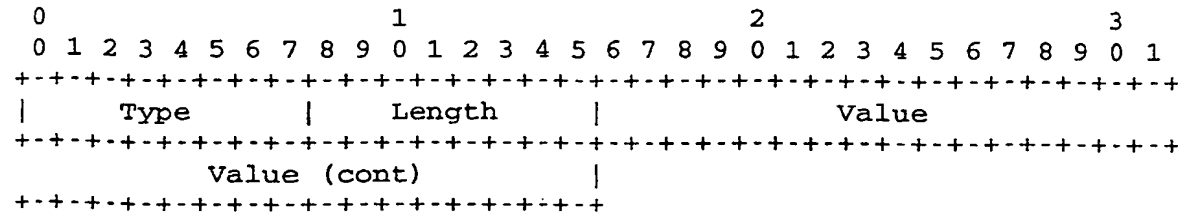
The Value field is four octets.

5.9. Acct-Output-Packets

Description

This attribute indicates how many packets have been sent to the port in the course of delivering this service to a Framed User, and can only be present in Accounting-Request records where the Acct-Status-Type is set to Stop.

A summary of the Acct-Output-Packets attribute format is shown below. The fields are transmitted from left to right.



Type

48 for Acct-Output-Packets.

Length

6

Value

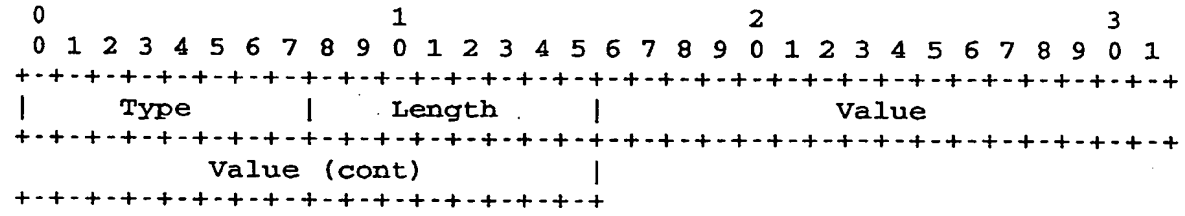
The Value field is four octets.

5.10. Acct-Terminate-Cause

Description

This attribute indicates how the session was terminated, and can only be present in Accounting-Request records where the Acct-Status-Type is set to Stop.

A summary of the Acct-Terminate-Cause attribute format is shown below. The fields are transmitted from left to right.



Type

49 for Acct-Terminate-Cause

Length

6

Value

The Value field is four octets, containing an integer specifying the cause of session termination, as follows:

1 User Request
2 Lost Carrier
3 Lost Service
4 Idle Timeout
5 Session Timeout
6 Admin Reset
7 Admin Reboot
8 Port Error
9 NAS Error
10 NAS Request
11 NAS Reboot
12 Port Unneeded
13 Port Preempted
14 Port Suspended
15 Service Unavailable
16 Callback
17 User Error
18 Host Request

The termination causes are as follows:

User Request	User requested termination of service, for example with LCP Terminate or by logging out.
Lost Carrier	DCD was dropped on the port.
Lost Service	Service can no longer be provided; for example, user's connection to a host was interrupted.
Idle Timeout	Idle timer expired.
Session Timeout	Maximum session length timer expired.
Admin Reset	Administrator reset the port or session.

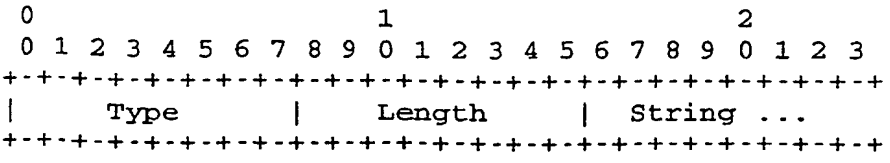
Admin Reboot	Administrator is ending service on the NAS, for example prior to rebooting the NAS.
Port Error	NAS detected an error on the port which required ending the session.
NAS Error	NAS detected some error (other than on the port) which required ending the session.
NAS Request	NAS ended session for a non-error reason not otherwise listed here.
NAS Reboot	The NAS ended the session in order to reboot non-administratively ("crash").
Port Unneeded	NAS ended session because resource usage fell below low-water mark (for example, if a bandwidth-on-demand algorithm decided that the port was no longer needed).
Port Preempted	NAS ended session in order to allocate the port to a higher priority use.
Port Suspended	NAS ended session to suspend a virtual session.
Service Unavailable	NAS was unable to provide requested service.
Callback	NAS is terminating current session in order to perform callback for a new session.
User Error	Input from user is in error, causing termination of session.
Host Request	Login Host terminated session normally.

5.11. Acct-Multi-Session-Id

Description

This attribute is a unique Accounting ID to make it easy to link together multiple related sessions in a log file. Each session linked together would have a unique Acct-Session-Id but the same Acct-Multi-Session-Id. It is strongly recommended that the Acct-Multi-Session-Id contain UTF-8 encoded 10646 [7] characters.

A summary of the Acct-Session-Id attribute format is shown below. The fields are transmitted from left to right.



Type

50 for Acct-Multi-Session-Id.

Length

>= 3

String

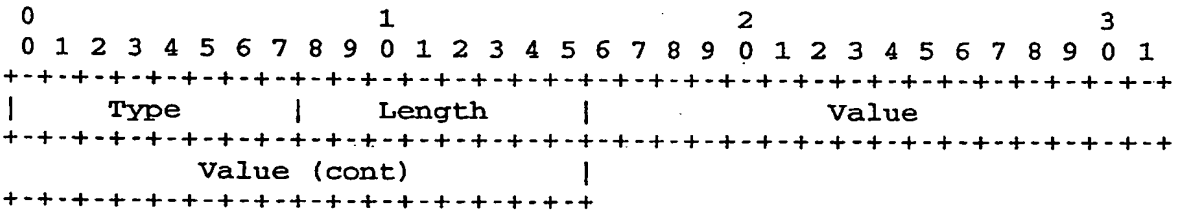
The String field SHOULD contain UTF-8 encoded 10646 [7] characters.

5.12. Acct-Link-Count

Description

This attribute gives the count of links which are known to have been in a given multilink session at the time the accounting record is generated. The NAS MAY include the Acct-Link-Count attribute in any Accounting-Request which might have multiple links.

A summary of the Acct-Link-Count attribute format is show below. The fields are transmitted from left to right.



Type

51 for Acct-Link-Count.

Length

6.

Value

The Value field is four octets, and contains the number of links seen so far in this Multilink Session.

It may be used to make it easier for an accounting server to know when it has all the records for a given Multilink session. When the number of Accounting-Requests received with Acct-Status-Type = Stop and the same Acct-Multi-Session-Id and unique Acct-Session-Id's equals the largest value of Acct-Link-Count seen in those Accounting-Requests, all Stop Accounting-Requests for that Multilink Session have been received.

An example showing 8 Accounting-Requests should make things clearer. For clarity only the relevant attributes are shown, but additional attributes containing accounting information will also be present in the Accounting-Request.

Multi-Session-Id	Session-Id	Status-Type	Link-Count
"10"	"10"	Start	1
"10"	"11"	Start	2
"10"	"11"	Stop	2
"10"	"12"	Start	3
"10"	"13"	Start	4
"10"	"12"	Stop	4
"10"	"13"	Stop	4
"10"	"10"	Stop	4

5.13. Table of Attributes

The following table provides a guide to which attributes may be found in Accounting-Request packets. No attributes should be found in Accounting-Response packets except Proxy-State and possibly Vendor-Specific.

#	Attribute
0-1	User-Name
0	User-Password
0	CHAP-Password

0-1	NAS-IP-Address [Note 1]
0-1	NAS-Port
0-1	Service-Type
0-1	Framed-Protocol
0-1	Framed-IP-Address
0-1	Framed-IP-Netmask
0-1	Framed-Routing
0+	Filter-Id
0-1	Framed-MTU
0+	Framed-Compression
0+	Login-IP-Host
0-1	Login-Service
0-1	Login-TCP-Port
0	Reply-Message
0-1	Callback-Number
0-1	Callback-Id
0+	Framed-Route
0-1	Framed-IPX-Network
0	State
0+	Class
0+	Vendor-Specific
0-1	Session-Timeout
0-1	Idle-Timeout
0-1	Termination-Action
0-1	Called-Station-Id
0-1	Calling-Station-Id
0-1	NAS-Identifier [Note 1]
0+	Proxy-State
0-1	Login-LAT-Service
0-1	Login-LAT-Node
0-1	Login-LAT-Group
0-1	Framed-AppleTalk-Link
0-1	Framed-AppleTalk-Network
0-1	Framed-AppleTalk-Zone
1	Acct-Status-Type
0-1	Acct-Delay-Time
0-1	Acct-Input-Octets
0-1	Acct-Output-Octets
1	Acct-Session-Id
0-1	Acct-Authentic
0-1	Acct-Session-Time
0-1	Acct-Input-Packets
0-1	Acct-Output-Packets
0-1	Acct-Terminate-Cause
0+	Acct-Multi-Session-Id
0+	Acct-Link-Count
0	CHAP-Challenge

0-1	NAS-Port-Type
0-1	Port-Limit
0-1	Login-LAT-Port

[Note 1] An Accounting-Request MUST contain either a NAS-IP-Address or a NAS-Identifier (or both).

The following table defines the above table entries.

0	This attribute MUST NOT be present
0+	Zero or more instances of this attribute MAY be present.
0-1	Zero or one instance of this attribute MAY be present.
1	Exactly one instance of this attribute MUST be present.

6. IANA Considerations

The Packet Type Codes, Attribute Types, and Attribute Values defined in this document are registered by the Internet Assigned Numbers Authority (IANA) from the RADIUS name spaces as described in the "IANA Considerations" section of RFC 2865 [2], in accordance with BCP 26 [8].

7. Security Considerations

Security issues are discussed in sections concerning the authenticator included in accounting requests and responses, using a shared secret which is never sent over the network.

8. Change Log

US-ASCII replaced by UTF-8.

Added notes on Proxy.

Framed-IP-Address should contain the actual IP address of the user.

If Acct-Session-ID was sent in an access-request, it must be used in the accounting-request for that session.

New values added to Acct-Status-Type.

Added an IANA Considerations section.

Updated references.

Text strings identified as a subset of string, to clarify use of UTF-8.

9. References

[1] Rigney, C., "RADIUS Accounting", RFC 2139, April 1997.

[2] Rigney, C., Willens, S., Rubens, A. and W. Simpson, "Remote Authentication Dial In User Service (RADIUS)", RFC 2865, June 2000.

[3] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March, 1997.

[4] Postel, J., "User Datagram Protocol", STD 6, RFC 768, August 1980.

[5] Rivest, R. and S. Dusse, "The MD5 Message-Digest Algorithm", RFC 1321, April 1992.

[6] Reynolds, J. and J. Postel, "Assigned Numbers", STD 2, RFC 1700, October 1994.

[7] Yergeau, F., "UTF-8, a transformation format of ISO 10646", RFC 2279, January 1998.

[8] Alvestrand, H. and T. Narten, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 2434, October 1998.

10. Acknowledgements

RADIUS and RADIUS Accounting were originally developed by Steve Willens of Livingston Enterprises for their PortMaster series of Network Access Servers.

11. Chair's Address

The RADIUS working group can be contacted via the current chair:

Carl Rigney
Livingston Enterprises
4464 Willow Road
Pleasanton, California 94588

Phone: +1 925 737 2100
EMail: cdr@telemancy.com

12. Author's Address

Questions about this memo can also be directed to:

Carl Rigney
Livingston Enterprises
4464 Willow Road
Pleasanton, California 94588

EMail: cdr@telemancy.com

13. Full Copyright Statement

Acknowledgement

Funding for the RFC Editor function is currently provided by the Internet Society.

Network Working Group
Request for Comments: 1272

C. Mills
BBN
D. Hirsh
Meridian Technology Corporation
G. Ruth
BBN
November 1991

INTERNET ACCOUNTING: BACKGROUND

Status of this Memo

This memo provides information for the Internet community. It does not specify an Internet standard. Distribution of this memo is unlimited.

1. Statement of Purpose

This document provides background information for the "Internet Accounting Architecture" and is the first of a three document set:

Internet Accounting Background & Status	(this document)
Internet Accounting Architecture	(under construction)
Internet Accounting Meter Service	(under construction)

The focus at this time is on defining METER SERVICES and USAGE REPORTING which provide basic semantics for measuring network utilization, a syntax, and a data reporting protocol. The intent is to produce a set of standards that is of practical use for early experimentation with usage reporting as an internet accounting mechanism.

The architecture should be expandable as additional experience is gained. The short-term Internet Accounting solution is intended to merge with OSI and Autonomous Network Research Group (ANRG) efforts and be superseded by those efforts in the long term. The OSI accounting working groups are currently defining meter syntax and reporting protocols. The ANRG research group is currently researching economic models and accounting tools for the Internet environment.

Internet Accounting as described here does not wrestle with the applications of usage reporting, such as monitoring and enforcing network policy; nor does it recommend approaches to billing or tackle such thorny issues as who pays for packet retransmission.

This document provides background and tutorial information on issues

surrounding the architecture, or in a sense, an explanation of choices made in the Internet Accounting Architecture.

2. Goals for a Usage Reporting Architecture

We have adopted the accounting framework and terminology used by OSI (ISO 7498-4 OSI Reference Model Part 4: Management Framework). This framework defines a generalized accounting management activity which includes calculations, usage reporting to users and providers and enforcing various limits on the use of resources. Our own ambitions are considerably more modest in that we are defining an architecture to be used over the short- term (until ISO and ANRG have final pronouncement and standards) that is limited to network USAGE REPORTING.

The OSI accounting model defines three basic entities:

1) the METER, which performs measurements and aggregates the results of those measurements;

2) the COLLECTOR, which is responsible for the integrity and security of METER data in short-term storage and transit; and

3) the APPLICATION, which processes/formats/stores METER data. APPLICATIONS implicitly manage METERS.

This working group, then, is concerned with specifying the attributes of METERS and COLLECTORS, with little concern at this time for APPLICATIONS.

3. The Usage Reporting Function

3.1. Motivation for Usage Reporting

The dominant motivations for usage reporting are:

- Understanding/Influencing Behavior.
  Usage reporting provides feedback for the subscriber on his use of network resources. The subscriber can better understand his network behavior and measure the impact of modifications made to improve performance or reduce costs.

- Measuring Policy Compliance.
  From the perspective of the network provider, usage reports might show whether or not a subscriber is in compliance with the stated policies for quantity of

network usage. Reporting alone is not sufficient to enforce compliance with policies, but reports can indicate whether it is necessary to develop additional methods of enforcement.

- Rational Cost Allocation/Recovery.
  Economic discipline can be used to penalize inefficient network configuration/utilization as well as to reward the efficient. It can be used to encourage bulk transfer at off hours. It can be used as a means to allocate operating costs in a zero-sum budget, and even be used as the basis for billing in a profit-making fee-for-service operation.

The chief deterrent to usage reporting is the cost of measuring usage, which includes:

- Reporting/collection overhead.
  This offers an additional source of computational load and network traffic due to the counting operations, managing the reporting system, collecting the reported data, and storing the resulting counts. Overhead increases with the accuracy and reliability of the accounting data.

- Post-processing overhead.
  Resources are required to maintain the post-processing tasks of maintaining the accounting database, generating reports, and, if appropriate, distributing bills, collecting revenue, servicing subscribers.

- Security overhead.
  The use of security mechanisms will increase the overall cost of accounting. Since accounting collects detailed information about subscriber behavior on the network and since these counts may also represent a flow of money, it is necessary to have mechanisms to protect accounting information from unauthorized disclosure or manipulation.

The balance between cost and benefit is regulated by the GRANULARITY of accounting information collected. This balance is policy-dependent. To minimize costs and maximize benefit, accounting detail is limited to the minimum amount to provide the necessary information for the research and implementation of a particular policy.

3.2. Network Policy and Usage Reporting

Accounting requirements are driven by policy. Conversely, policy is typically influenced by the available management/reporting tools and their cost. This section is NOT a recommendation for billing practices, but intended to provide additional background for understanding the problems involved in implementing a simple, adequate usage reporting system.

Since there are few tools adequate for any form of cost recovery and/or long-term monitoring there are few organizations that practice proactive usage reporting in the Internet. Those that do have generally invented their own. But far and away the most common approach is to treat the cost of network operations as overhead with network reports limited to short-term, diagnostic intervention. But as the population and use of the Internet increases and diversifies, the complexity of paying for that usage also increases. Subsidies and funding mechanisms appropriate to non-profit organizations often restrict commercial use or require that "for profit" use be identified and billed separately from the non-profit use. Tax regulations may require verification of network connection or usage. Some portions of the Internet are distinctly "private", whereas other Internet segments are treated as public, shared infrastructure.

The number of administrations operating in some connection with the Internet is exploding. The network "hierarchy" (backbone, regional, enterprise, stub network) is becoming deeper (more levels), increasingly enmeshed (more cross-connections) and more diversified (different charters and usage patterns). Each of these administrations has different policies and by-laws about who may use an individual network, who pays for it, and how the payment is determined. Also, each administration balances the OVERHEAD costs of accounting (metering, reporting, billing, collecting) against the benefits of identifying usage and allocating costs.

Some members of the Internet community are concerned that the introduction of usage reporting will encourage new billing policies which are detrimental to the current Internet infrastructure (though it is also reasonable to assert that the current lack of usage reporting may be detrimental as well). Caution and experimentation must be the watch words as usage reporting is introduced. Well before meters are used for active BILLING and ENFORCEMENT, they should first be used to:

- UNDERSTAND USER BEHAVIOR
  (learn to quantify and/or predict individual and aggregate traffic patterns over the long term),

- QUANTIFY NETWORK IMPROVEMENTS,
  (measure user and vendor efficiency in how network resources are consumed to provide end-user data transport service) and

- MEASURE COMPLIANCE WITH POLICY.

Accounting policies for network traffic already exist. But they are usually based on network parameters which change seldom, if at all. Such parameters require little monitoring (the line speed of a physical connection, e.g.,Ethernet, 9600 baud, FDDI). The connection to the network is then charged to the subscriber as a FLAT-FEE regardless of the amount of traffic passed across the connection and is similar to the monthly unlimited local service phone bill.

Usage-insensitive access charges are sufficient in many cases, and can be preferable to usage-based charging in Internet environments, for financial, technical, and social reasons. Sample incentives for the FLAT-FEE billing approach are:

- FINANCIAL:
  Predictable monthly charges. No overhead costs for counting packets and preparing usage-based reports.

- TECHNICAL:
  Easing the sharing of resources. Eliminating the headaches of needing another layer of accounting in proxy servers which associate their usage with their clients'. Examples of proxy servers which generate network traffic on behalf of the actual user or subscriber are mail daemons, network file servers, and print spoolers.

- SOCIAL:
  Treating the network as an unregulated public infrastructure with equal access and information sharing. Encouraging public-spirited behavior -- contributing to public mailing lists, information distribution, etc.

In other cases USAGE-SENSITIVE charges may be preferred or required by a local administration's policy. Government regulations or the wishes of subscribers with low or intermittent traffic patterns may force the issue (note: FLAT FEES are beneficial for heavy network users. USAGE SENSITVE charges generally benefit the low-volume user). Where usage-sensitive accounting is used, cost ceilings and floors may still be established by static parameters, such as "pipe size" for fixed connections or "connection time" for dial-up connection, to satisfy the need for some predictability.

Different billing schemes may be employed depending on network measures of distance. For example, local network traffic may be flat-rate and remote internet traffic may be usage-based, analogous to the local and long distance billing policies adopted by the telephone companies.

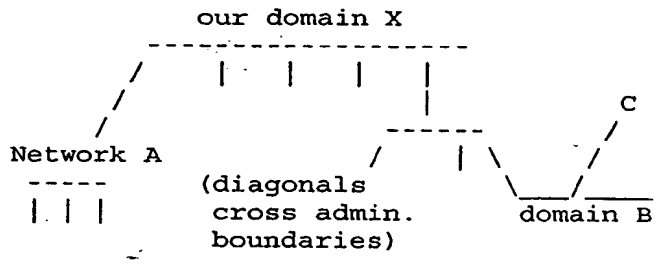
The ANRG is independently investigating policy models and infrastructure economics for billing and cost recovery.

3.3. The Nature of Usage Accounting

Although the exact requirements for internet usage accounting will vary from one network administration to the next and will depend on policies and cost trade-offs, it is possible to characterize the problem in some broad terms and thereby bound it. Rather than try to solve the problem in exhaustive generality (providing for every imaginable set of accounting requirements), some assumptions about usage accounting are posited in order to make the problem tractable and to render implementations feasible. Since these assumptions form the basis for our architectural and design work, it is important to make them explicit from the outset and hold them up to the scrutiny of the Internet community.

3.3.1. A Model for Internet Accounting

We begin with the assumption that there is a "network administrator" or "network administration" to whom internet accounting is of interest. He "owns" and operates some subset of the internet (one or more connected networks)that may be called his "administrative domain". This administrative domain has well defined boundaries.



The network administrator is interested in 1) traffic within his boundaries and 2) traffic crossing his boundaries. Within his boundaries he may be interested in end-system to end-system accounting or accounting at coarser granularities (e.g., department to department).

The network administrator is usually not interested in accounting for end-systems outside his administrative domain; his primary concern is accounting to the level of other ADJACENT (directly connected) administrative domains. Consider the viewpoint of the administrator for domain X of the internet. The idea is that he will send each adjacent administrative domain a bill (or other statement of accounting) for its use of his resources and it will send him a bill for his use of its resources. When he receives an aggregate bill from Network A, if he wishes to allocate the charges to end users or subsystems within his domain, it is HIS responsibility to collect accounting data about how they used the resources of Network A. If the "user" is in fact another administrative domain, B, (on whose behalf X was using A's resources) the administrator for X just sends his counterpart in B a bill for the part of X's bill attributable to B's usage. If B was passing traffic for C, them B bills C for the appropriate portion X's charges, and so on, until the charges percolate back to the original end user, say G. Thus, the administrator for X does not have to account for G's usage; he only has to account for the usage of the administrative domains directly adjacent to himself.

This paradigm of recursive accounting may, of course, be used WITHIN an administrative domain that is (logically) comprised of sub-administrative domains.

The discussion of the preceding paragraphs applies to a general mesh topology, in which any Internet constituent domain may act as a service provider for any connected domain. Although the Internet topology is in fact such a mesh, there is a general hierarchy to its structure and hierarchical routing (when implemented) will make it logically hierarchical as far as traffic flow is concerned. This logical hierarchy permits a simplification of the usage accounting perspective.

At the bottom of the service hierarchy a service-consuming host sits on one of many "stub" networks. These are interconnected into an enterprise-wide extended LAN, which in turn receives Internet service, typically from a single attachment to a regional backbone. Regional backbones receive national transport services from national backbones such as NSFnet, Alternet, PSInet, CERFnet, NSInet, or Nordunet. In this scheme each level in the hierarchy has a constituency, a group for which usage reporting is germane, in the level underneath it. In the case of the NSFnet the natural constituency, for accounting purposes at least, is the regional nets (MIDnet, SURAnet,...). For the regionals it will be their member institutions; for the institutions, their stub networks; and for the stubs, their individual hosts.

3.3.2. Implications of the Model

The significance of the model sketched above is that Internet accounting must be able to support accounting for adjacent (intermediate) systems, as well as end-system accounting. Adjacent system accounting information cannot be derived from end-system accounting (even if complete end-system accounting were feasible) because traffic from an end-system may reach the administrative domain of interest through different adjacent domains, and it is the adjacent domain through which it passes that is of interest.

The need to support accounting for adjacent intermediate systems means that internet accounting will require information not present in internet protocol headers (these headers contain source and destination addresses of end-systems only). This information may come from lower layer protocols (network or link layer) or from configuration information for boundary components (e.g., "what system is connected to port 5 of this IP router").

4. Meters

A METER is a process which examines a stream of packets on a communications medium or between a pair of media. The meter records aggregate counts of packets belonging to FLOWs between communicating entities (hosts/processes or aggregations of communicating hosts (domains)). The assignment of packets to flows may be done by executing a series of rules. Meters can reasonably be implemented in any of three environments -- dedicated monitors, in routers or in general-purpose systems.

Meter location is a critical decision in internet accounting. An important criterion for selecting meter location is cost, i.e., REDUCING ACCOUNTING OVERHEAD and MINIMIZING THE COST OF IMPLEMENTATION.

In the trade-off between overhead (cost of accounting) and detail, ACCURACY and RELIABILITY play a decisive role. Full accuracy and reliability for accounting purposes require that EVERY packet must be examined. However, if the requirement for accuracy and reliability is relaxed, statistical sampling may be more practical and sufficiently accurate, and DETAILED ACCOUNTING is not required at all. Accuracy and reliability requirements may be less stringent when the purpose of usage-reporting is solely to understand network behavior, for network design and performance tuning, or when usage reporting is used to approximate cost allocations to users as a percentage of total fees.

Overhead costs are minimized by accounting at the coarsest acceptable

GRANULARITY, i.e., using the greatest amount of AGGREGATION possible to limit the number of accounting records generated, their size, and the frequency with which they are transmitted across the network or otherwise stored.

The other cost factor lies in implementation. Implementation will necessitate the development and introduction of hardware and software components into the internet. It is important to design an architecture that tends to minimize the cost of these new components.

4.1. Meter Placement

In the model developed above, the Internet may be viewed as a hierarchical system of service providers and their corresponding constituencies. In this scheme the service provider accounts for the activity of the constituents or service consumers. Meters should be placed to allow for optimal data collection for the relevant constituency and technology. Meters are most needed at administrative boundaries and data collected such that service provider and consumer are able to reconcile their activities.

Routers (and/or bridges) are by definition and design placed (topologically) at these boundaries and so it follows that the most generally convenient place to position accounting meters is in or near the router. But again this depends on the underlying transport. Whenever the service-providing network is broadcast (e.g., bus-based), not extended (i.e., without bridging or routing), then meter placement is of no particular consequence. If one were generating usage reports for a stub LAN, meters could reasonably be placed in a router, a dedicated monitor, or a host at any point on the LAN. Where an enterprise-wide network is a LAN, the same observation holds. At the boundary between an enterprise and a regional network, however, in or near a router is an appropriate location for meters that will measure the enterprise's network activity.

Meters are placed in (or near) routers to count packets at the Internet Protocol Level. All traffic flows through two natural metering points: hosts and routers (Internet packet switches). Hosts are the ultimate source and sink of all traffic. Routers monitor all traffic which passes IN or OUT of each network. Motivations for selecting the routers as the metering points are:

- Minimization of cost and overhead.
  (by concentrating the accounting function). Centralize and minimize in terms of number of geographical or administrative regions, number of protocols monitored, and number of separate implementations modified. (Hosts are too diverse and numerous for easy standardization.

Routers concentrate traffic and are more homogeneous.)

- Traffic control.
  When and if usage sensitive quotas are involved, changes in meter status (e.g., exceeding a quota) would result in an active influence on network traffic (the router starts denying access). A passive measuring device cannot control network access in response to detecting state.

- Intermediate system accounting.
  As discussed above, internet accounting includes both end-system and intermediate system accounting. Hosts see only end-system traffic; routers see both the end-systems (internet source and destination) and the adjacent intermediate systems.

Therefore, meters should be placed at:

- administrative boundaries
  only for measuring inter-domain traffic;

- stub networks
  for measuring intra-domain traffic. For intra-domain traffic, the requirement for performing accounting at almost every router is a disincentive for implementing a usage-based charging policy.

4.2. Meter Types

Four possible types of metering technology are:

- Network monitors:
  These measure only traffic WITHIN a single network. They include LAN monitors, X.25 call accounting systems and traffic monitors in bridges.

- Line monitors:
  These count packets flowing across a circuit. They would be placed on inter-router trunks and on router ports.

- Router-integral meters:
  These are meters located within a router, implemented in software. They count packets flowing through the router.

- Router spiders:
  This is a set of line monitors that surround a router, measure traffic on all of its ports and coordinate the results.

4.3. Meter Structure

While topology argues in favor of meters in routers, granularity and security favor dedicated monitors. The GRANULARITY of the accountable entity (and its attributes) affects the amount of overhead incurred for accounting. Each entity/attribute/reporting interval combination is a separate meter. Each individual meter takes up local memory and requires additional memory or network resources when the meter reports to the application. Memory is a limited resource, and there are cost implications to expanding memory significantly or increasing the frequency of reporting. The number of concurrent flows open in a router is controlled by

- the granularity of the accountable entity
- the granularity of the attributes and sub-categories of packets
- memory
  (the number of flows that can be stored concurrently, a limit which can also be expressed as the average number of flows existing at this granularity plus some delta, e.g., peak hour average plus one standard deviation, or ...)
- the reporting interval
  (the lifetime of an individual meter)

There is a spectrum of granularity control which ranges across the following dimensions. (Most administrations will probably choose a granularity somewhere in the middle of the spectrum.)

ENTITY: Entities range across the spectrum from the coarsest granularity, PORT (a local view with a unique designation for the subscriber port through which packets enter and exit "my" network) through NETWORK and HOST to USER (not defined here). The port is the minimum granularity of accounting. HOST is the finest granularity defined here. Where verification is required, a network should be able to perform accounting at the granularity its subscribers use. Hosts are ultimately responsible for identifying the end user, since only the hosts have unambiguous access to user identification. This information could be shared with the network, but it is the host's responsibility to do so, and there is no mechanism in place at this time (e.g., an IP option, discussed in section 4.).

ATTRIBUTE: Each new attribute requires that an additional flow be maintained for each entity. The coarsest granularity is NO

categorization of packets. The finest granularity would be to maintain state information about the higher-levels protocols or type of service being used by communicating processes across the network.

VALUES: Values are the information which is recorded for each entity/attribute grouping. Usually values are counters, such as packet counts and byte counts. They may also be time stamps - start time and stop time, or reasons for starting or stopping reporting.

REPORTING INTERVAL: At the very finest level of granularity, each data packet might generate a separate accounting record. To report traffic at this level of detail would require approximately one packet of accounting information for every data packet sent. The reporting interval is then zero and no memory will be needed for flow record storage. For a non-zero reporting interval flow records must be maintained in memory. Storage for stale (old, infrequent) flows may be recycled when their data has been reported. As the reporting interval increases, more and more stale records accumulate.

The feasibility of a particular group of granularities varies with the PERFORMANCE characteristics of the network (link speed, link bandwidth, router processing speed, router memory), as well as the COST of accounting balanced against the requirement for DETAIL. Since technological advances can quickly obsolete current technical limitations, and since the policy structure and economics of the Internet are in flux, meters will be defined with VARYING GRANULARITY which is regulated according to the traffic requirements of the individual network or administration and technical limitations.

4.4. Collection Issues

There are two implicit assumptions about the nature of meters and traffic sources that they measure, both of which have substantial bearing on collectors.

1. The matrix of communicating entity pairs is large but sparse and, moreover, network traffic exhibits considerable source, destination and attribute coherence - so that lists can be quite compact.

2. Meters can be configured to generate either a static set of variables whose values are incremented, or a stream of records that must be periodically transferred and removed from the meter's memory.

Meters can generate large, unstructured amounts of information and the essential collection issue revolves around mapping collection activities into an SNMP framework (or, to the extent that this is not successful, specifying other collection paradigms).

There are three major collection concerns:

- o data confidentiality
- o data integrity
- o local and remote collection control

The prime security concern is preserving the confidentiality of usage data. (See ISO 7498 Part 2, "Security Architecture," for security terminology used herein.) Given that accounting data are sensitive, the collector should be able (or may be required) to provide confidentiality for accounting data at the point of collection, through transmission and up to the point where the data is delivered. The delivery function may also require authentication of the origin and destination and provision for connection integrity (if connections are utilized). Other security services (e.g., measures to counter denial of service attacks) are not deemed necessary for internet accounting at this time. It is assumed that security services can be provided by SNMP and its mechanisms. (This will require further investigation.)

In order to have an accurate monitoring system, reliable delivery of data should be assured through one or more of:

- o an acknowledgement retransmission scheme;
- o redundant reporting to multiple collectors;
- o having backup storage located at the meter.

There is a place for both application polling and meter traps within this scheme, but there are significant trade-offs associated with each.

Polling means that the collection point has some control over when accounting data is sent, so that not all meters flood the collector at once. However, polling messages, particularly when structured with SNMP's GET-NEXT operator, add considerable overhead to the network. Meter traps are required in any case (whether or not polling is the preferred collection method), so that a meter may rid itself of data when its cache is full.

The fundamental collection trade-off will be between primary and secondary storage at the meter, coupled with an efficient bulk-transfer protocol, versus minimal storage at the meter and a network-bandwidth-consuming collection discipline.

A final collection concern is whether packets should be counted on entry into a router or upon exit from a router. It is the nature of IP that not every packet received by a router is actually passed to an output port. The Internet Protocol allows routers to discard packets (e.g., in times of congestion when the router cannot handle the offered load); it is presumed that higher level protocols (e.g., TCP) will provide whatever reliable delivery service the user deems necessary (by detecting non- delivery and retransmitting).

The question arises, therefore, whether an internet accounting system should count all packets offered to a router (since each packet offered consumes some router resources) or just those that are finally passed by the router to a network (why should a user pay for undelivered packets?) Since there are good arguments for either position, we do not attempt to resolve this issue here. (It should be noted, however, that SMDS has chosen to count on exit only.) Rather, we require that an internet accounting should provide ability for counting packets either way -- on entry to or on exit from a router.

5. Examples

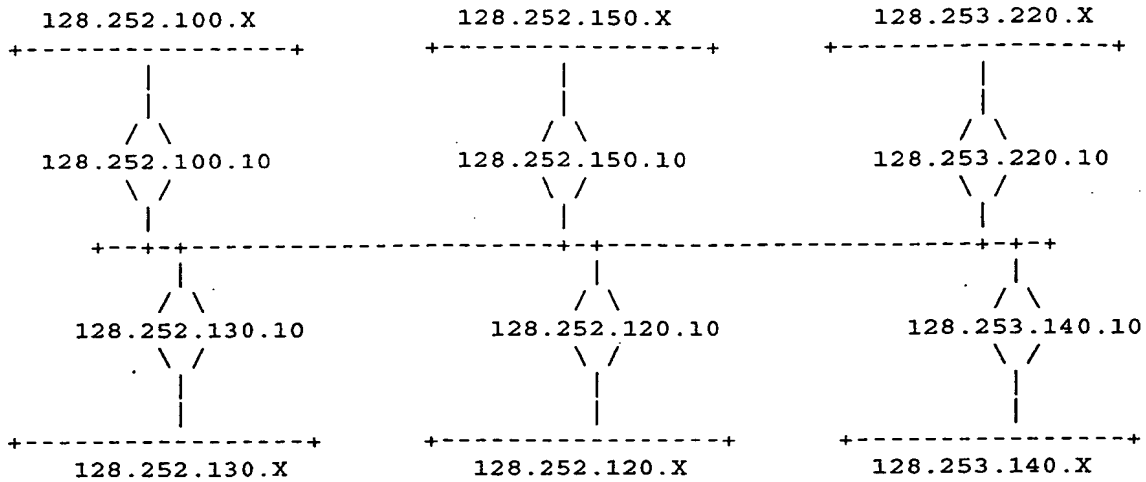
Here follows a series of examples to illustrate what data may be of interest to service providers and consumers in a number of different scenarios. In the illustrations that follow straight lines are interpreted as some sort of LAN. Diagonals are point- to-point links. Diamonds are routers. We assume that we are in a homogeneous protocol environment (IP).

5.1 A Single Segment LAN

Consumers and providers on a single LAN service can utilize the same set of data: the contribution of individual hosts to total network load. A network accounting system measures flows between individual host pairs. (On a broadcast LAN, e.g., an Ethernet, this can be accomplished by a single meter placed anywhere on the LAN.) Using this data, costs for the network management activity can be apportioned to individual hosts or the departments that own/manage the hosts.

Alternately, flows can be kept by source only, rather than source-destination pairs.

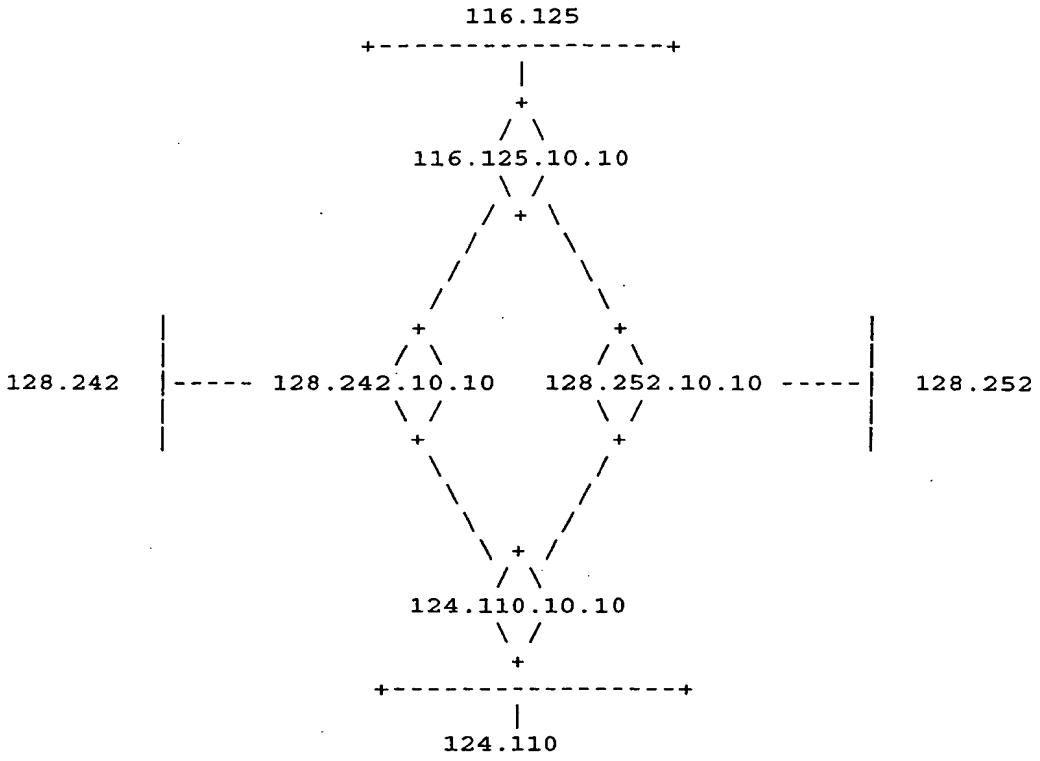
5.2 An Extended (Campus or Facility-Wide) LAN



This is the first example in which the information that is germane for service provider and consumer are not identical. The service consumers are now the individual subnets and the service provider is the facility-wide backbone. A service provider is interested in knowing the contribution of individual subnets to the total traffic of the backbone. In order to ascertain this, a meter on the backbone (the longest line in the center of the illustration) can keep track of flows between subnet pairs. Now the communications between individual hosts on adjacent subnets are aggregated into a single flow that measures activity between subnets.

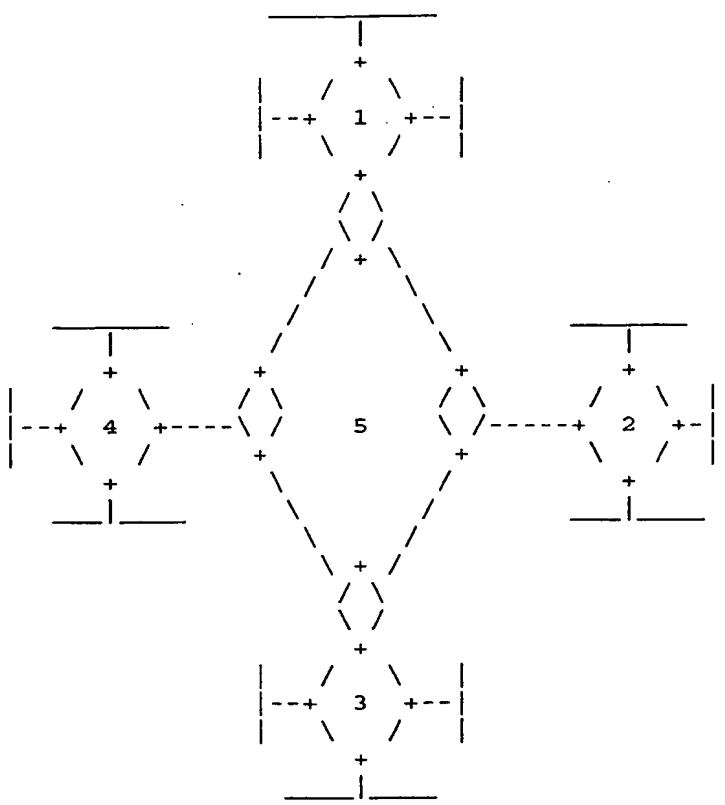
The service consumers, or subnets, might in turn want to keep track of the communications between individual hosts that use the services of the backbone. An accounting system on the backbone could be configured to monitor traffic among individual host pairs. Alternately an accounting system on each individual subnet could keep track of local and "non-local" traffic. The observed data of the two sets of meters (one for the service provider and one for the service consumers) should have reconcilable data.

5.3 A Regional Network



In this example we have a regional network consisting of a ring of point-to-point links that interconnect a collection of campus-wide LANs. Again service provider and consumer have differing interests and needs for accounting data. The service provider, the regional network, again will be interested in the contribution of each individual network to the total traffic on the regional network. This interest might extend to include measure of individual link utilization, and not just total offered load to the network as a whole. In this latter case the service provider will require that meters be placed at one end or the other on each link. For the service consumer, the individual campus, relevant measures would include the contribution of individual subnets or hosts to the total "outbound" traffic. Meter(s) placed in (or at) the router that connects the campus- network to the regional network can perform the necessary measurement.

5.4 A National Backbone



In this last case, the data that the service provider will want to collect is the traffic between regional networks. The flow that measures a regional network, or regional network pairs, is defined as the union of all member-campus network address spaces. This can be arrived at by keeping multiple individual network address flows and developing the regional network contribution as post-processing activity, or by defining a flow that is the union of all the relevant addresses. (This is a cpu cycles for memory trade-off.) Note that if the service provider measures individual network contributions, then this data is, in large measure, the data that the service consumers would require.

6. Future Issues

This last section is the collector for ancillary issues that are as yet undefined or out of current scope.

APPLICATIONS standards: Recommendations for storage, processing and reporting are left out for the moment. Storage and processing of accounting information is dependent on individual network policy. Recommendations for standardizing billing schemes would be premature.

QUOTAS are a form of closed loop feedback that represent an interesting extension of usage reporting. But they will have to wait until the basic accounting technology is reasonably defined and has been the subject of a reasonable amount of experimentation.

SESSION ACCOUNTING: Detailed auditing of individual sessions across the internet (at level four or higher) will not be addressed by internet accounting. Internet accounting deals only with measuring traffic at the IP level.

APPLICATION LEVEL ACCOUNTING: Service hosts and proxy agents have to do their own accounting for services, since the network cannot distinguish on whose behalf they are acting. Alternately, TCP/UDP port numbers could become an optional field in a meter, since the conjunction of a pair of IP addresses and port numbers occurring at a particular time uniquely identifies a pair of communicating processes.

The USER has not yet been defined, since an IP option would have to be added to the IP header to provide for this. This option would probably contain two parts - a subscriber identification and a user sub-identification - to allow for the later introduction of quota mechanisms which have both group and individual quotas. The subscriber is the fiscally responsible entity, for example the manager of a research group. In any case, routers must be able to fall back to accounting by host, since there will most certainly be hosts on the network which do not implement a new IP option in a timely fashion.

7. References

International Standards Organization (ISO), "Management Framework," Part 4 of Information Processing Systems Open Systems Interconnection Basic Reference Model,ISO 7498-4, 1984.

International Standards Organization (ISO), "Security Architecture," Part 2 of Information Processing Systems Open Systems Interconnection Basic Reference Model,ISO 7498-2, 1984.

Security Considerations

Security issues are discussed in sections 2, 3 and 4.

Authors' Addresses

Cyndi Mills
Bolt, Beranek, and Newman
150 Cambridge Park Drive
Cambridge, MA 02140

Phone: 617-873-4143
Email: cmills@bbn.com

Donald Hirsh
Meridian Technology Corporation
11 McBride Corporate Center Drive
Suite 250
Chesterfield, MO 63005

Phone: 314-532-7708
Email: hirsh@meridian.uucp

Gregory Ruth
Bolt, Beranek, and Newman
150 Cambridge Park Drive
Cambridge, MA 02140

Phone: 617-873-3150
Email: gruth@bbn.com

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G06F 13/00, 13/10	A1	(11) International Publication Number: WO 97/22936
		(43) International Publication Date: 26 June 1997 (26.06.97)

(21) International Application Number: PCT/US96/19689

(22) International Filing Date: 12 December 1996 (12.12.96)

(30) Priority Data:
08/574,528 19 December 1995 (19.12.95) US

(71) Applicant: MOTOROLA INC. [US/US]; 1303 East Algonquin Road, Schaumburg, IL 60196 (US).

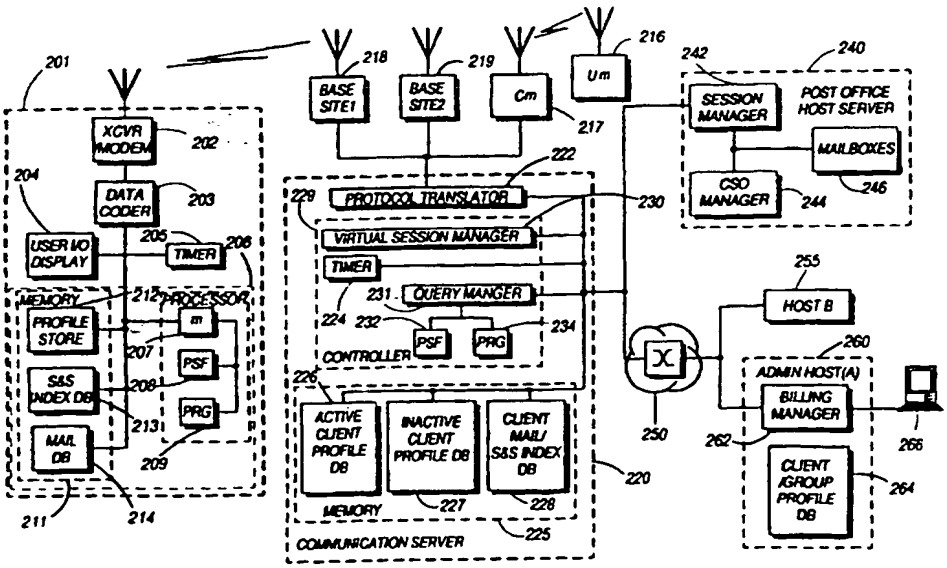
(72) Inventors: EGGLESTON, Gene; 1303 Mink Trail, Cary, IL 60013 (US). HANSEN, Mitch; 241 Foxmoor, Fox River Grove, IL 60021 (US). KREBS, Richard; 1201 Cougar Trail, Cary, IL 60013 (US).

(74) Agents: WOOD, J., Ray et al.; Motorola Inc., Intellectual Property Dept., 1303 East Algonquin Road, Schaumburg, IL 60196 (US).

(81) Designated States: CA, CN, GB.

Published
With international search report.

(54) Title: METHOD AND APPARATUS FOR RATE GOVERNING COMMUNICATIONS



(57) Abstract

A system including a rate governor is provided for monitoring and controlling the amount of communications between a remote communication unit (201) and communication server (220). Preferably, as thresholds are passed, a user is alerted to amounts (time and/or charges) spent or remaining, and once a use limit is reached, further communication is restricted. A main rate governor (234) is maintained at the communication server (220), allowing access, control and the like by administrators (260) and the like. A further rate governor (209), responsive to the main rate governor, may also be used at the remote unit (201). By means of the rate governors a method is provided for both limiting user or group data transfer beyond a set amount, as well as providing alerts to users as a limit is approached.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgystan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic			SE	Sweden
CG	Congo	KR	Republic of Korea	SG	Singapore
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LR	Liberia	SZ	Swaziland
CS	Czechoslovakia	LT	Lithuania	TD	Chad
CZ	Czech Republic	LU	Luxembourg	TG	Togo
DE	Germany	LV	Latvia	TJ	Tajikistan
DK	Denmark	MC	Monaco	TT	Trinidad and Tobago
EE	Estonia	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	UG	Uganda
FI	Finland	ML	Mali	US	United States of America
FR	France	MN	Mongolia	UZ	Uzbekistan
GA	Gabon	MR	Mauritania	VN	Viet Nam

METHOD AND APPARATUS FOR RATE GOVERNING COMMUNICATIONS

Field Of The Invention

The present invention relates to communications and more particularly an improved method and apparatus for transferring data in a communications system.

Background

The last 10 years have seen a tremendous increase in the demand for communications services, including both wired and wireless networks capable of handling data communications. Unlike real-time voice services, such as standard telephony or cellular wireless services, in which circuit-switched communications are used because of the sensitivity of users to the timing of oral dialogue/voice data, greater efficiencies can often be achieved in non-voice data communications through the use of packet-switched or hybrid communications systems. This is particularly the case with communications to remote users (e.g., persons sending messages via one of the well-known available wireless networks like GSM (Global System for Mobiles) or AMPS (Advanced Mobile Phone System) cellular), where protracted circuit-switched sessions into a mail server or LAN (local area network) could be prohibitively expensive due to the high per-minute session charges by the wireless service provider.

One solution to this problem has been for users to limit, as much as feasible, their communications to sessionless

communications. This can be done, e.g., by subscribing to additional email services that can receive LAN/WAN (wide area network) email and send out broadcast pages and transmissions to registered users, in lieu of requiring a user to maintain a session with a mail server. However, this disadvantageously requires subscription to an additional service, and is typically limited in the types of applications supported. With the rapid growth in emerging session-oriented applications--like the popular client server application of Lotus Notes®--the need is growing for more cost effective solutions to providing connectivity of such session-oriented applications and users remotely located from their host servers.

Regardless of whether a session-oriented or session-less communication service is used, it is also desirable to limit the amount of information communicated between a remote user and host, both to save off-site user's time and to limit the costs arising from the more expensive rates for remote communications. Unfortunately, typical applications like email do not provide for user-selected methods for choosing and limiting the volume of down-loaded communications, or for filtering uploaded or downloaded communications. Thus, a user who wants to receive remote messaging is left with an option of receiving all his messages (or some summary thereof), even the ones he or she might otherwise be willing to leave unprocessed until a later time when no longer using expensive remote communications services. Further, many processes, like that of a typical email reply, are wasteful of bandwidth by resending all earlier messages each time a new reply is generated, even though the earlier messages may still be stored at both ends of the wireless network.

In addition to the above concerns over how to optimize the types and amount of data being transferred, there is additionally a problem in a lack of effective techniques for monitoring and even controlling an aggregate use of tariffed networks. While the network service providers have means for tracking use by an individual unit

basis, which is totaled in periodic billing statements, this information is typically unavailable to users or their managers/application administrators. Thus, users and managers are typically left without any effective means for controlling the level of messaging during a billing cycle, and can only monitor or react to usage following the service providers periodic statements.

There remains therefore a need for an improved means for data communications that solves these and related problems.

Brief Description Of The Drawings

FIG. 1 is a block diagram of a communications system according to a first embodiment of the invention;

FIG. 2 is a block diagram of a communications system according to a further embodiment of the invention;

FIG. 3 is a flow chart illustrating virtual session data transfer between the different functional entities of the wireless communications system of FIG. 2;

FIG. 4 is a flow chart illustrating a pre-stage filtering embodiment for data transfer between the different functional entities of the wireless communications system of FIG. 2;

FIG. 5 is a flow chart illustrating one embodiment of pre-stage filtering for data transfers;

FIG. 6 is a flow chart illustrating another embodiment of pre-stage filtering for data transfers;

FIG. 7 is a flow chart illustrating a message summarization and selection embodiment for data transfer between the different functional entities of the wireless communications system of FIG. 2;

FIG. 8 is a diagram illustrating an embodiment of a summary index for use in the process of FIG. 7;

FIG. 9 is a flow chart illustrating an optimized reply embodiment for data transfer between the different functional entities of the wireless communications system of FIG. 2; and

FIG. 10 is a flow chart illustrating a rate governor embodiment for data transfer between the different functional entities of the wireless communications system of FIG. 2.

Detailed Description

These problems and others are solved by the improved method and apparatus according to the invention. A presently preferred first main embodiment of the invention is a system including a virtual session manager (VSM) for establishing and maintaining a sessionless communication path with a first data processing device (e.g., a mobile client) on the one hand and a session-oriented communication path with a second data processing device (e.g., a host system). The session-oriented communication protocol (including network and application layer protocols) with the host system permits remote access to, e.g., LAN-based applications, while the virtual session, via a sessionless-oriented communication protocol, between the VSM and remote (i.e., coupled via a tariffed network or connection) client permits this access to be carried out without the expense of a dedicated/circuit switched connection.

In a second main embodiment, a prestage filter stage is provided for applying user-definable filter parameters (e.g., reject, pass, or granularity filters) on data being transferred between the remote communication unit and communication server. For downloading, e.g., email from a host post office, a communication server controller preferably either forwards the filter parameters in a query object or message to the post office to apply and return

qualified mail, or the communication server receives all unprocessed mail and applies the filters locally, only acknowledging as processed that mail which is qualified. For uploading, e.g., email from a client, a client controller applies an upload prestage filter so as to retain all filter rejected mail, while transmitting mail passing the filters. Thus, only desired data transfers (i.e., those meeting user defined filters) are communicated over the expense-bearing networks between the remote unit and communication server.

In yet another main embodiment, a select and summary (S&S) listing or index is used to provide user flexibility in reviewing and requesting otherwise filtered data. Both the user's remote communication unit and communication server maintain a S&S index containing identifying (summary) information about data which has not been fully transferred between the communication unit and communication server. As new data is reviewed and filtered for transfer, identifying/summary information is captured for any non-qualifying data by either a host unit or the communication server. This information is stored in the communication server's S&S index, and at least periodically, or upon request, transferred via update messaging to the remote communication unit. Upon reviewing its updates or its S&S index, the user may send a request for such of the data that it desires partial or full transfers for further review. Thus, a cost efficient review mechanism is provided to users for determining whether to transfer data that otherwise fails selected filter parameters.

In a fourth main embodiment, a method and apparatus for optimized reply to messaging is provided. When sending a reply, the remote communication unit's controller generates a delta (e.g., data representing the content difference between two messages) between a preceding message and the reply message, and forms an optimized reply using the delta and an identifier of the preceding message. On receiving the optimized reply, the communication server uses the data unit identifier to retrieve the preceding message from a further

host (e.g., the post office mailbox of the user associated with the remote unit), reconstructs the full reply from the retrieved message and the delta, and forwards the full reply to the addressee. When receiving a reply for the remote unit, an index is preferably maintained at both the remote unit and communication server of mail stored at the remote unit. Resort is made to this index to determine a preceding message forming part of the reply. An optimized reply is similarly formed from a delta and identifying information of the preceding message, and sent to the remote unit. In this manner, the volume and expense incurred in reply messaging is greatly reduced, by only sending a delta and small header (i.e., the identifying information).

Finally, in a fifth embodiment, a rate governor is provided for monitoring and controlling the amount of communications between the remote unit and communication server. Preferably, as threshold(s) are passed a user is alerted to amounts (time and/or charges) spent or remaining, and once a use limit is reached further communication is restricted. A main rate governor is maintained at the communication server, allowing access, control and the like by system administrators and the like. A further rate governor, responsive to the main rate governor, may also be used at the remote unit. By means of this rate governor a mechanism is provided for both limiting user or group data transfer beyond a set amount, as well as providing alerts to users as the limit is approached.

Turning now to FIG. 1, there is generally depicted a communication system 100 in accordance with a first embodiment of the invention. This system is configured to support one or more user devices such as wireless subscriber units (i.e., mobile station (MS) 105) communicating with host processor 115 via an infrastructure including base station 120 and intermediate system 125 coupled to a data network 130. In the illustrated case mobile station 105 is a portable computer having an rf (radio frequency) modem 106. A communications server 110, including a virtual session manager

(VSM) and query manager (QM), is coupled between the public data network 130 and the host server 115. The virtual session manager and query manager are, preferably, an appropriately configured data processing device, the VSM and QM program being shipped for loading on the server 110 via any convenient means such as a machine-readable CD-ROM 111 (compact disc-read only memory) or the like. Counterpart client-communications software, e.g., a prestage filter, can be shipped via a similar convenient form like CD-ROM 107, downloaded directly from server 110 to subscriber 105 (also being, e.g., a data processing device, by which is meant virtually any processor (but not a human) capable of processing data for a programmed result, whether a general purpose computer or more specialized electronic processor), or the like.

In this embodiment the mobile user 105 communicates with the server/VSM 110 using any appropriate data protocol being used by the data network 130, as necessarily modified for transport over the wireless infrastructure; the wireless infrastructure could be, e.g., any private system like ARDIS® or DataTAC®, CDPD (cellular digital packet data), GPRS (GSM Packet Radio Service), and the like. Thus, a sessionless data flow between the mobile user 105 and server/VSM 110 occurs on an event driven basis, and no costly connection is maintained when there is nothing being communicated. In order to keep connectivity costs to a minimum, the server 110 is preferably connected to the LAN/WAN on which the host 115 is also connected, via any standard LAN/WAN communication channel (e.g., a bus or backbone). This allows the communications server 110 to advantageously maintain the same session with the host 115 that the client 105 typically enjoys when connected to the LAN/WAN. Thus, by use of the server 110 the client 105 can achieve a virtual session with the host 115 with almost the same access as if directly connected to the host's 115 LAN, but at a substantial reduction in the cost of communicating via the wireless network and PDN 130.

FIG. 2 illustrates an alternative communication system 200 embodiment of the present invention. A first client, a mobile end system (M-ES) computer including a user device 201, is in communication with a base station (BS1) 218 of a wireless communication system. This base station 218 is coupled, e.g., on a same bus or via bridges/routers, to a communication server 220 which includes VSM 230. An electronic mail (email) post office is coupled locally to VSM 230, either as another program running on the same communications server 220 or located on another server 240 of the communications server's 220 LAN/WAN. It is not important, however, where the post office is located for purposes of operation of the VSM 230, as is illustrated by other application hosts B and C 255, 260 being in communication via other networks such as a public data network or public switched telephone network 250. In fact, the same user 201 could be concurrently coupled via the VSM 230 to, for example, a local email post office 240, a remote client-server host 255, a further database host server (not shown), an administrator host server 260, a multimedia host, a voice processor, etc. It should be understood that for purposes of this application, a first device or component is responsive to or in communication with a second unit or component regardless of whether the first and second units are directly coupled or indirectly coupled, such as via intermediate units, including switches that operatively couple the units for only a segment of time, as long as a signal path can be found that directly or indirectly establishes a relationship between the first and second units. For example, the client computer 105 is in communication with the VSM server 110 even though intermediate system (e.g., a router or switch) 125 and a packet network 130 having multiple switches etc. are disposed between the user device 105 and VSM server 110.

In the illustrated case client 201 includes a data transfer manager or exchange unit 206, which in simple form could be an appropriately programmed electronic processor 207 (e.g., a general purpose CPU (central processing unit) and memory or data store 211.

A timer 205 is also preferably employed in the data exchange control process, as will be explained further in connection with the flow chart of FIG. 3 below. A typical client 201 would also include some form(s) of user interface such as display 204, a data encoder/decoder 203 to accommodate the system communications protocol(s), and a transceiver (if using rf or infrared communications) and a modulator-demodulator (or modem) 202 to connect to a wireless or wireline communications network. Transceiver/modem 202 in this case would either include a built-in or attached user module for wireless LAN communications; the specific type will vary depending on the system, e.g., including PCMCIA (personal computer memory card interface association) wireless modems, and attached or built-in PSTN (public switched telephone network) modem, etc. Specific features of data exchange unit 206 preferably includes (as more fully described below) a prestage filter (PSF) manager 208, rate governor (RG) 209, user profile store 212, select and summary index store 213, and mail store 214 (a store being any available device (e.g., ROM (read-only memory), disks) or program (e.g., a database) for storage of the specified information).

The communication server 220 preferably includes a data transfer manager or controller 229 having a VSM 230, memory stores for storing active client profile (user parameters) and inactive client profile information 226 and 227, a timer 224, and optionally some form of protocol translators or formatters 222. The VSM 230 serves to manage the virtual session with the client 201 and session with host systems 240, 255 and/or 260 based on the parameters loaded into the active user parameter store/profile memory 226 or object. Controller 229 preferably also includes a query manager (QM) 231 for controlling specific processes (e.g., sending messages to a post office to query for unprocessed messages and forwarding received messages etc.), and a prestage filter 232 and rate governor 234. Memory 225 also preferably includes a client select and summary index database or store 228, which will also be described more fully

below in connection with FIGS. 7 and 8. The protocol translators 222 serve to format or code the messages as appropriate for transport between the VSM 230 and client 201; these include, e.g., appropriate protocol software that can be located at the communications server, or any other convenient processor per design of the given communication system. By messages is meant any appropriate data unit (whether a frame, datastream, packet, or other format), including objects, datagrams, etc., for containing information being communicated.

Communications server 220 is also illustrated as supporting additional users, e.g. user module 216, communicating via different access points, e.g., control module (CM) 217 of a wireless LAN and base station 219, all access points 217-219 being coupled via a common bus, backbone, etc. These base stations can be part of the same communication system, similar systems owned by different service providers, or even different systems, all of which may be different from the communications server service provider. Thus, for example, a single communications server can support at one local region 215 an ARDIS® node, a RAM® node, a wireless LAN controller module, a CDPD node, an in-building cordless telephone node, etc., allowing users from a variety of systems to access the same communications server and post office. Users not registered could access through the appropriate one of these nodes along the model of FIG. 1, i.e., via PDN 250 to a remote communications server having their VSM/QM. Thus, any number of system configurations is possible, limited only by the network services provided and the user's preference.

A process by which a VSM manages communications between client and host is illustrated in the flow chart embodiment of FIG. 3. This process typically begins with a user event, such as instantiation (forming) of a communications object at the client and sending a registration message (steps 301-302). Alternatively, the infrastructure could initiate the communications by sending a page

or the like requesting the client to register (for example, when the client has registered with the wireless system but not yet requested registration with the communications server). In any event, once a registration message is received by the communications server, it preferably authenticates and otherwise qualifies the client, including sending a logon/registration message to the host for its authentication of the client (steps 303-305). Upon successful authentication, the communications server instantiates a client object (CO) for the communications session including client parameters retrieved from an inactive client parameter store, as modified by the user in his registration or subsequent messages (step 306). These parameters include at a minimum client and host identifiers, but may also include additional preferences based on the type of communications involved. Also, the registration and authentication process can be handled by the VSM, or alternatively by another appropriately programmed entity of the communications server. Following instantiation at the server, a response message, e.g., a further registration message, is sent to the client, and an acknowledgment (ACK) returned to the server; both client and server then retain the instantiated objects as fully qualified, and may start session timers (steps 307-309). At this point a virtual session has been established between the client and the VSM, and a regular session established between the VSM and host computer. If the registration is not successful, then any instantiated object is deleted, with the client returned to an inactive status.

Upon establishing the virtual session, a query is preferably generated by query manager requesting unprocessed data for the user, and the VSM forwards the query to the host (step 320). In the case of email, e.g., this might include generating a request message for all unread mail in the users post office box. The post office then checks for new mail received, and forwards all such mail to the VSM (steps 321-322). Because the VSM has established a LAN session with the post office, these communications are performed relatively quickly, e.g., in accordance with the LAN's and host's typical processing for

their current loading level. The VSM in turn forwards the data (i.e., mail) received via the virtual session transport (step 323). For example, in the case of FIG. 1 where PDN 130 is an ISDN (integrated services digital network) network connected to a CDPD wireless network, the mail would be appropriately packetized by the communications server and delivered via the serving BS 120 according to ISDN/CDPD system protocols. This can take up to several minutes or more for a moderately sized mail package. However, since the data is being delivered in a sessionless mode, the amount of time the communication channel (including the more expensive wireless communication channel portion, as well as the portion via PDN 130) is tied up is kept to a minimum. This also translates into a significant cost savings for the user, since the user is only charged on a per packet basis for mail when it is actually transported, and doesn't have to pay for a prolonged session to keep connected to the post office in order to receive new mail. Finally, upon receipt by the client, appropriate acknowledgments are sent and the post office box updated, e.g., by marking the mail as read or processed (steps 324-326)

While in some systems it may be advantageous to store some of the data at the communications server, in the case of email and the like it is presently envisioned that the communication server is preferably used in maintaining the sessions between client and host, and not as a remote server for the host. Thus, rather than have all new data from the host pushed down to the communications server, most data exchanges are preferably initiated, at some predetermined interval or intervals, by the communications server (e.g., by the query manager).

Further, it is an inefficient use of resources to continue querying a host or attempting to deliver data when the client is no longer receiving at its remote location (occurring, e.g., when the client leaves a coverage area, or the user turns off its modem or processor). Thus, a process for either maintaining the client in an

active status, or removing the client from active status in response to an event, is also preferably included in the VSM. One such process is to utilize timers at both client and VSM to determine when a virtual session is no longer active. The timers are first set upon registration, and are subsequently reset after each data exchange (steps 327-336). If no data exchange occurs within a predetermined period of time, say 20 minutes, both client and VSM would remove the client qualification (i.e., destroy the client object for the communication session) and, if desired, mark the client as being in an inactive status (steps 337-340). The VSM would also forward a logoff message to the host (step 341). In order to avoid an undesired time out, the client is preferably configured to send a short message after a predetermined period since the last data exchange, sufficiently prior to the time at which the timers elapse so that the VSM can receive it. Otherwise, if there are only intermittent data exchanges, the client may be required to frequently re-register; this in turn means the client will not be notified of outbound data until the client re-registers and is again coupled via the virtual session manager.

Turning now to FIGS. 4 through 6, a presently preferred embodiment is shown for prestage filtering data for transfer between the different functional entities of the wireless communications system of FIG. 2. This typically begins with the generation of a query object or message at the communications server (step 406). This object/message may be created in response to a preceding client generated message (e.g., a request generated when clicking on an application button requesting updates, executing the mail application, etc.), or in response to settings in the client profile. However, after updating the active client profile/object for an active client application, the query manager is preferably programmed to send query objects at predetermined intervals for each application being run by each active client, the intervals varying depending on the application type or administrator preference (e.g., for mail about every 10-30 seconds or longer). Alternatively, the

intervals could be user specified via the client profile, for example to shorten the query intervals for time critical applications (e.g., for emergency services or "real time" applications) , or lengthen the intervals when less frequent updates are desired (e.g., to conserve on traffic expenses for updates to a rapidly changing, but non-time critical, group-ware file or document).

The content of the query objects will vary depending both upon the application and client filter settings. One approach for mail applications is to have a predetermined number of user-definable filter attributes stored in the client profile databases (e.g., stores 212 and 226-227 of FIG. 2). These attributes can include, by way of example, the priority of a message (e.g., urgent, normal, or low); the date on which the message is sent or posted; the size of the message (typically uncompressed, i.e., the normal stored size; although transmission size or cost could also be used); the author of the message; and the subject of the message (e.g., key words in a subject line or in the text). These attributes can simply be used as reject criteria (e.g., reject all messages having "low" priority, date before "12/15/95", size more than "2" kbytes (kilobytes), or subject not containing "project x"), pass criteria (all messages from "Boss") or a combination of both, the variety and complexity being a matter of design choice. These attributes also preferably include certain "granularity" filters, i.e., filters additionally limiting the size of a message passing all or most of the other filters. Three possible examples of granularity filters are a truncation size filter (e.g., truncate the message after the first "100" bytes), and text or file attachment filters (e.g., indicating whether or not to strip attachments). Thus, messages passing all criteria but message size could still be received in a truncated size meeting the message size criterion. Alternatively, messages failing the author or subject filters could still be passed with header information, by setting all rejected messages to be passed with a text truncation size of "0" bytes. One skilled in the art will appreciate that a variety of other reject/pass filter criteria may be used, and the specific ones and

combinations of user-definable (or even administrator-definable) features will be largely a matter of design choice depending on factors such as the desired functionality, complexity, and application(s) (including filterable features). It is significant, however, that clients are now provided by the present invention with a means for effecting prestage filtering of their communications by virtue of the communications server and definable filter settings, rather than having to choose between receiving no messages or receive all messages, including less important or expensive and time-consuming transmissions.

The prestage filtering is preferably performed at the host server. This may be accomplished, for example, by passing the filter attributes in an appropriately formatted query object or message for use by the host application. In the illustrated case a query object with the client filter settings is forwarded to the post office, and applied by a communications server object or CSO (instantiated at the post office when the virtual session is established). The post office/CSO reads/queries the query object for the filter attributes, and applies these criteria in the selection and formatting of unprocessed messages (steps 408-412). The filtered messages are then encapsulated and forwarded to the QM, which similarly forwards the filtered messages (with appropriate protocol translation) to the client (steps 414-416). Alternatively, where the host application is not designed to permit prestage filtering, all unprocessed messages can be forwarded to the communications server, where the filters are applied via a prestage filter (PSF) object or routine (e.g., PSF 232 of FIG. 2), with only qualifying/filtered messages being forwarded to the client (steps 410, 418-424). Through acknowledgments the post office is notified how to mark the mail index in both cases. For example, when prestage filtering at the post office, all forwarded mail would be marked as processed/read and all filtered mail as unprocessed (truncated messages being marked as either depending on design conventions, or if available marked as filtered or partially processed). If prestage filtering is done at the communications

server only those messages forwarded to the client would be acknowledged and marked as processed (step 428).

In addition to download/downlink filtering, prestage filtering is also advantageously used in upload/uplink transmissions. This can take the form of granularity filtering, or automatically retaining the whole data unit or message based upon filterable attributes for later transmission when on a lower cost network. In this case, each client would have a prestage filter (PSF) unit such as that of PSF 208 of FIG. 2 (e.g., a PSF object or routine drawing on selected attributes in the profile store 212). Each data unit generated is filtered using the user-selected criteria, with qualifying data being forwarded via the communication server (steps 430-436). If a data unit is not sent, it is retained locally for transmission later, e.g., when connected via a lower cost network to the post office. As an enhancement, the user could additionally be provided with a selection of types of send buttons (i.e., filtered send or unfiltered send), or be prompted with an alert dialogue or similar message when a message is filtered to decide whether to forward the data unfiltered (steps 438-440). Similarly, the user can be provided with several groups of filter settings that could be manually or automatically activated, so as to enable the client to adjust plural filter settings with a minimum effort, for example by switching to a more restrictive profile when entering important meetings (which profile could be automatically activated via an appropriately configured and coupled calendar program, etc.).

While only the client need retain the upload filter attributes in its profile store, preferably both the communication server and client store copies of the download filter settings in their profile memories. This conveniently permits a client to review all settings whenever desired, and to change the settings locally. When the download settings are changed at the client, the changes are communicated to the communication server preferably as soon as the change is made, or as soon as a virtual session is established if the

changes are made while offline from the communication server (steps 442-444). Further, where a summary index of filtered messages is maintained (as is described in connection with FIGS. 7 and 8 below), upon a change in filter settings the communication server may be automatically set to forward all messages previously rejected but now passing the new filter settings.

FIGS. 5 and 6 illustrate two approaches to prestage filtering particularly useful for email filtering. In FIG. 5, a series of five reject filters are applied to each message. If a mail message does not meet any of the criteria (priority, date, size, author, or subject/key word) then it is left unprocessed (steps 502-516). Once all unreviewed messages (i.e., all unprocessed messages, or if expanded marking is available all unprocessed messages not previously filtered) have been filtered, those not rejected are forwarded (step 518). FIG. 6 illustrates the application of granularity filters. If a message exceeds the filter size, it is appropriately truncated (including insertion of a note indicating truncation) (steps 602-606). Similarly, if there are text or file attachments, and these are marked to be filtered, they are stripped with, optionally, a note being inserted alerting the addressee that the attachment was stripped (steps 608-614). Once filtered, the message is sent (step 616).

FIGS. 7 and 8 illustrate a further enhancement, permitting the user to more conveniently review selected information even for filtered/rejected data. In the preferred embodiment a query object or message is similarly generated by the communication server as described above. However, in addition to the profile information, the query object in this case includes a request for summary information about each partially and fully rejected message (step 702). When the host (i.e., a post office server in the illustrated case) receives the query it applies the appropriate filters; if only qualifying mail is present, this is forwarded to the client as described above (steps 704-708). Where there is partially (e.g., truncated) or fully rejected

data, identifying summary information is captured for all rejected data (step 710). For mail this identifying summary information would include the message serial number, along with certain header information (801 and 802 of FIG. 8). This header information may include any filterable attribute (e.g., date, author, subject, size, priority, attachment indicator) and is preferably client definable, so the client can decide how much header information it needs and how much to omit. All qualifying and non-qualifying (i.e., filter-rejected) mail is marked similarly as described above (step 712).

When the response object or message is received by the QM of the communication server, the encapsulated identifying summary information is saved to a select and summary (S&S) index, such as that illustrated by client S&S index database 228 of FIG. 2 and the index structure of FIG. 8. This index is preferably created in response to the first query following full qualification, although one could retain a stored index when the client is inactive as long as the index is fully updated upon re-registration/ qualification. In order to minimize transmissions between the communication server and the client, only changes to the S&S index are forwarded, as summary delta data (i.e., a delta of the revised index to the immediately preceding index, the preceding index being an acknowledged version same as that stored in the S&S index (e.g., S&S index database 213 of FIG. 2) of the client). Where only identifying summary information is received in response to the query object, one may additionally delay forwarding the delta information to the client for a predetermined period of time or until the next message passing the prestage filters is forwarded, whichever comes first (i.e., the filter-rejected information more likely being less important, some users may prefer to receive S&S index updates less frequently in order to further reduce costs or interruptions) (steps 714-718).

Upon receiving the delta of the identifying summary information, the client updates its S&S index and, when appropriate, prompts the user (again, the prompt criteria could be set for all

messages, or some sub-set based on any filterable attribute, etc.). The user is thus able to review the summary information and make a determination on whether or not to override the filter rejection. For mail the user wants to read, the user indicates the decision by any appropriate means (clicking on the message, voice command, etc.) and an appropriate request generated (e.g., for all selected mail, for only a partially filtered version (e.g., truncated), etc.) (steps 720-722). The request is appropriately translated, as needed, and sent as a query object or message to the post office. Upon retrieval, the requested data is forwarded to the client via the QM. Upon receipt at the client, a read acknowledgment may be generated and sent to the communication server. Preferably when the read acknowledgment is received at the communication server a further ACK (acknowledgment signal) may be sent to the client, at which time both client and communication server update their respective S&S indices to remove the entry for read mail from the S&S index, and note any partially read mail. Upon acknowledgment, the post office may further mark any read mail as processed (steps 724-734).

As with prestage filtering, one skilled in the art will appreciate that many more filterable attributes and summary inputs are possible than those described, and which ones are available will depend on such factors as the desired functionality, complexity, and application(s) (including filterable features) for which the select and summary index is being used. The index structure may thus similarly vary significantly, as will the means for achieving similar indices for both the client and communication server; in other words, while one could simply periodically forward the whole index, where practical any one of a number of known delta (e.g., data representing the content difference between two files) or other update approaches for communicating less than the whole index are likely more preferable. What is significant, no matter the particular design approach selected, is that a summary index, showing unprocessed or partially processed data (e.g., that filtered), is available to a client for determination on whether to process the data further, with a

substantially identical index being retained at the communication server in order to further reduce transmission requirements.

FIG. 9 illustrates a yet further improvement, this embodiment permitting a user to minimize the data transmitted for responses to earlier data transmissions. This is particularly advantageous in the case of email, where it is common to append all prior messages in an email conversation to a reply, making for lengthy reply messages that contain substantial portions that are identical to mail already saved at the client or target unit. While this has come to be expected in email replies, it is also quite costly in time and tariff charges in bandwidth limited systems like most wireless communication systems.

Starting from a client perspective, the process of FIG. 9 commences with a client formulating a reply to a received mail message, much as he or she would for any typical email application (step 902). However, when the user executes the reply, e.g., by clicking on a send button, the client controller (201 of FIG. 2) optimizes the reply message by calculating a delta or difference, using any appropriate delta routine, between the reply message and the preceding message. This delta is then formed into an optimized reply along with a message/data unit identifier for the preceding message/data unit (preferably the mail serial number, although any retrievable identifier of the preceding message may be used, such as header information, or even a CRC (cyclic redundancy check) value) (step 904). To ensure that only the shortest message is being sent, the controller additionally compares the reply message with the optimized reply to determine which is optimum for transmission (step 906). This determination may be made based on a comparison of the message sizes, compressed and formatted message sizes, or any other convenient means for estimating which version of the reply will require the least bandwidth or transport cost. Thus, for example, a normal reply message to a very short message may be selected for transmission where the overhead of the delta and

message identifier make the optimized reply bigger than the normal reply message would be. However, in most instances it is anticipated that the optimized reply will be smaller than a normal reply message, providing significant savings to the client in time and costs.

When the optimized reply is received at the QM of the communication server, a determination is made on whether to reconstruct the normal reply message (i.e., form a replica reply) or to forward the optimized reply, based on known parameters (if any) of the target communication unit/client. Thus, for example, where both the originating and target clients are active and served by the same communication server and thus are known to have optimized reply capabilities, and the target client was an addressee or originator of the preceding message identified by the message identifier of the optimized reply, a reconstructed reply may not be required. Rather, since the preceding message would either be in the inbox or outbox of the target unit, the target unit can reconstruct the reply message from the identified mail in its mailbox and the delta. This advantageously allows bandwidth to be minimized for both the sending and target clients. Further, if perchance the target unit has already deleted the identified preceding message, the controller of the target unit could, rather than acknowledge receipt, send a request for the normal reply message, which the communication server would reconstruct as described next.

In cases where the target unit is not an active client with the communication server, the QM (or other appropriate entity of the controller) functions to reconstruct the reply message from the optimized reply. Because the communication server preferably does not retain a copy of client mail or data located on other hosts (such remote stores typically adding complexity and cost, while being unnecessary in view of the virtual session established via the communication server), it would use the identifier to retrieve the preceding message from the host (e.g., send a query object or

message to the appropriate post office) (steps 908-912). This can be implemented by requesting the preceding message from the client inbox, or from the originating unit's outbox (or even the target unit's inbox, if it is a cc: on the preceding message). Because the serial number is a unique number widely used in email applications, this is the preferable message identifier for email systems. However, where this unique number is unavailable other identifiers may be used, including author, date and/or subject matches. Further, for some messages it may even be advantageous to use other relatively unique values, such as CRC or other values, by themselves or together with other identifiers. It is relatively unimportant for purposes of the invention what the identifier is, as long as it is useful within the accuracy demanded by the system design for retrieving the correct preceding message.

Once the preceding message has been received by the communication server, it uses a counterpart delta routine to that of the client to reconstruct a replica of the reply message from the delta of the optimized reply and the retrieved copy of the preceding message. Once reconstructed, the reply message is forwarded to the target unit(s), as well as to the outbox or sent mail folder of the client's post office box (steps 914-916). While some additional processing and network traffic is required between the communication server and host, this is relatively inexpensive compared to the savings achieved by using an optimized reply over the tariffed network between the communication server and client.

While the preceding approach can be implemented without resort to a message index, it can be further optimized by use of indices at the communication server and client. In this case, a full index of each active client's mailbox (or other application file(s)) is maintained at both the client and the communication server. This index could advantageously be one of the S&S indices 213 and 228 of FIG. 2 designed to include all mail (although perhaps with less identifying information for received mail than for filter-restricted

mail, depending on factors such as the memory available and the amount of identifying information desirable). When an optimized reply is received at the communication server, a search of the appropriate client index (e.g., first the target unit, if also an active client, otherwise the client's or originating unit's indices) for the message identifier of the preceding message, indicating whether or not the preceding message has been deleted. When the preceding message's identifier is present, the process continues as noted above, in other words by sending the optimized reply to the target unit, or reconstructing the reply message and forwarding it to the target unit.

Replies being sent to the client can similarly use an optimized reply to minimize messaging sizes. Thus, for example, where a reply is received by the communication server which has the client as an addressee, the communication server is capable of generating a delta between the reply message and a preceding message known to be stored in a mail database (e.g., memory 214 of FIG. 2) of the client. The preceding message is most easily identified if an additional identifier is included with the reply for ease of searching in the client's index. However, where such is not included, identifier's can be extracted from the text (e.g., author, date, recipient, subject) for comparison matching. Alternatively, a comparison of the text of the reply message can be used in determining the preceding message. For example, a series of preceding messages could be retrieved for textual comparison; or alternatively an identifying value for all or selected (e.g., sent) mail can be maintained (e.g., by calculating the text CRC value and storing it in the index), and a check of selected portions (e.g., all portions below insertions identifying preceding messages in the text) of the reply message text can then be performed. The latest or largest matching preceding message is the selected (which could be either a message sent to, or sent from, the client), so as to minimize the delta, and the delta calculated between the preceding message and the reply message. An optimized reply is then formed including the delta and preceding message identifier

recognizable by the client. This optimized reply is then forwarded, and reconstructed at the client into the reply message. In other words, the client retrieves from memory the message corresponding to the message identifier, and forms a replica of the reply message from the delta and message. Once acknowledged, both client and communication server indices are appropriately updated to reflect the mail transfer (steps 918-930).

This embodiment thus provides an efficient process for sending reply data between a client and the communication server, without requiring the costly transfer of earlier transmitted portions of the reply data.

In a final embodiment, a rate governor is provided so as to assist clients in maintaining their messaging and expenses within desired limits. Turning to FIG. 10, with reference also to FIG. 2, one embodiment of such a rate governor is illustrated. This rate governor operates to track the approximate time and/or expense for client use, which can be as simple as timing a circuit-switched connection, or where packet data is being sent, timing (or estimating based on size) the time and/or cost of transmitting the packet over the tariffed network(s). In estimating the transmission value (e.g., cost), a rate governor could better estimate actual costs by taking into account known pricing factors established by each network service provider (e.g., rates by time of day, by grade/ quality of service (QoS) for packets, by size or bandwidth desired, etc.). These values would be maintained for application by the rate governor (234 of FIG. 2) as each data unit is received to determine an estimated transmission value.

In the illustrated case of an email application, upon receiving a client-generated message the QM (or other appropriate controller entity of the communication server) passes the pertinent packet information or message parameter (e.g., the packet size from the header) to the rate governor, which in this case operates as a packet rate governor (or PRG). The PRG determines from the client object

(or profile store) the amount of use time and/or charge still available (or alternatively, the amount already used, and limits allowed), and compares the use time remaining (e.g., a previously authorized or allocated transmission value) against the value for the message parameter (step 954).

Preferably several limits are established, including one or more alert thresholds. These alert thresholds would serve to warn the client each time a certain threshold is passed in amounts of time/charge used or remaining, permitting the client to limit use as needed to stay within budget, or to seek a higher limit in advance of the point at which the use limit is reached. This use or transmission limit serves as the budgeted limit for data transfers. Unless a user is privileged, once the use limit is reached further communications/data transfers are restricted. In the simplest form, such transfers are restricted by alerting the client that the use limit has been reached, terminating the current session and preventing further sessions until additional use limit time/charge is authorized. Alternatively, certain messaging could still be permitted (based, e.g., on any filterable criteria--e.g., permitting messages to the administrator but not a further communication unit), but with reminders that routine messages will not be forwarded. This would advantageously allow critical messages, messages to an administrator (e.g., requesting additional authorization), etc., to still be transferred, although it does not prevent a user from running up excess charges for messaging to the communication server. A PRG may thus also be advantageously used in the client (e.g., PRG 209 of FIG. 2), signaled by the PRG of the communication server to automatically set certain prestage filters to restrict all but certain message transfers until a new use limit is provided. If a user were to bypass this client PRG and continue improper messaging, all further sessions could be terminated by the communication server with notification to the administrator and client.

If the user is privileged, data transfers would still continue despite the user limit having been exceeded. However, an alert would still preferably be sent to both the client and administrator, allowing the administrator to verify the privilege and reset the use limit if desirable, and the client to still be aware it has passed a targeted use amount (steps 956-968 and 980-984). In any event, after each data transfer the client object or store is updated to reflect the new estimated transaction total (e.g., time remaining, total expense, etc.) (step 958).

As mentioned above, if a user is not privileged it is preferable to allow the client an additional data transfer to an administrator requesting additional allocation of time/ charges. This request would be forwarded by the communication server to the administrator host, where it would be processed for approval. If approved, the administrator would notify the communication server to adjust the use limit by a specified amount. Alternatively, if there is no system administrator, but charges or debits are handled through a communication server service provider, the client may send any appropriate authorization for additional charge/debit to the communication server (e.g., by sending an encrypted account number and identifying information like a PIN (personal identification number). Once the charge or debit is processed and approved to the service provider's satisfaction, the amount of charge or debit would be used to adjust the use limit. A notification would also be forwarded to the client of the new use limit, with the client PRG being updated accordingly (steps 970-978).

In addition to updating the use limit in response to a user or administrator request, the rate governor can also be advantageously set to automatically update the use limits upon the occurrence of a predetermined update event. Thus, for example, where billing and budgeting is done on a monthly cycle, and the administrator has set rate governor preferences so as to automatically reset the use limit on the first day of the next billing cycle, the communication server

will automatically reset the client use limit at the specified time and in the specified amount (step 992).

Moreover, in order to achieve an even more accurate billing control, the communication server could be coupled with the tariffed network service provider(s) so as to receive periodic charge statements for client data traffic, as well as updates for tariff rates, etc. In order to take advantage of these statements, a billing index would be maintained for each client estimating use and charges for each data transfer. Upon receiving the periodic charge statement (e.g., forwarded once a day during an administrative window) the estimated use entries are replaced by the actual use and charges from the statement, and the client profile (and object, if active) is updated to reflect a corrected use limit, etc. The administrator is notified, and the client is notified upon the next transaction, of the updated amount. If desired, the client or administrator can request a download of the current billing index showing the most recent estimated and actual charges (steps 986-990).

Finally, one should appreciate that the above process is equally applicable to groups as well as to individual clients. Thus, the PRG can advantageously be used to set use limits for groups and supergroups of users, as well as for individual clients as described above. Thus, where one of the applications being used is groupware, as opposed to the email example described above, different groups can be assigned group use limits for groupware data transfers (while retaining individual use limits for separate email or data transfers, etc.). To avoid one or two users exhausting the group's authorized limit, individual use limits can still be set for each client, although with more flexibility, e.g., to draw on unused group time before requiring additional allocation from an administrator, to permit another user of the group to yield a portion of its individual use limit, etc. As should be apparent, many variations exist on how the rate governor is structured, depending on the applications being used, clients and groups operating, the interactivity with service

providers, complexity or simplicity desired, and many other related and unrelated factors.

One skilled in the art will appreciate that there are many variations that are possible for the present invention, only a limited number of which have been described in detail above. Thus, for example, while the embodiments above describe application to clients communicating in certain systems, one should appreciate that it has application to any communication system, wired or wireless, client-server, distributed or other networks, etc., in which the user is remote from a host. It can also be used with almost any application program or groups of programs (e.g., transferring database, wordprocessing, graphics, voice etc. files, executing programs and control messages, etc.), not just email or groupware. Moreover, while processor 206, controller 229, timers 205 and 224, data stores 211 and 225, and other circuits, are described in terms of specific logical/functional/circuitry relationships, one skilled in the art will appreciate that such may be implemented in a variety of ways, preferably by appropriately configured and programmed processors, ASICs (application specific integrated circuits), and DSPs (digital signal processors), but also by hardware components, some combination thereof, or even a distributed architecture with individual elements physically separated but cooperating to achieve the same functionality. Thus, it should be understood that the invention is not limited by the foregoing description of preferred embodiments, but embraces all such alterations, modifications, and variations in accordance with the spirit and scope of the appended claims.

We claim:

Claims

1. A system for controlling communications with a communication unit (201) comprising:

a communication server (220), in communication with the communication unit, characterized by a data transfer manager (229) operable for communicating data between the communication unit and a host server, the data transfer manager comprising a rate governor (234, 209) for estimating a transmission value for data communicated between the communication server and communication unit, and comparing the estimated transmission value with an allocated transmission value.

2. The system of claim 1, wherein the rate governor is further operable for preventing communication of further data between the communication server and communication unit when the estimated transmission value exceeds the allocated transmission value.

3. The system of claim 1, wherein the communication server is in communication with the communication unit via at least a first communication channel, the first communication channel including a charge-bearing channel portion of a first communication service provider different from a second service provider of the communication server, the allocated transmission value corresponding to a total remaining transmission value authorized for the communication unit; and the rate governor is further operable for alerting the

communication unit when the allocated transmission value becomes less than a first threshold.

4. The system of claim 1, wherein the data transfer manager further comprises a client profile store for storing the allocated transmission value, and the rate governor is further operable for updating the allocated transmission value, in response to a first data unit having an estimated first transmission value, by subtracting the estimated first transmission value from the allocated transmission value to obtain an updated allocated transmission value, and replacing the allocated transmission value in the client profile store with the updated allocated transmission value.

5. The system of claim 1, wherein the data transfer manager further comprises a client profile store for storing the allocated transmission value, the data comprises plural data units and the estimated transmission value comprises an estimated total transmission value for the plural data units, and the rate governor is further operable for updating the allocated transmission value, in response to a further data unit having an estimated further transmission value, by adding the further estimated transmission value and estimated total transmission value to form an updated estimated total transmission value, and determining if the updated estimated total transmission value is greater than the allocated transmission value.

6. The system of claim 5, further comprising an administrator host server in communication with the communication server, wherein the rate governor is further operable to alert the administrator host server when the updated estimated total transmission value is greater than the allocated transmission value, and if the communication unit is

not privileged, prohibit further communication with the communication unit except to predetermined addressees.

7. The system of claim 1, further comprising a host server in communication with the communication server, wherein the data transfer manager further comprises a virtual session manager adapted to control communication of data between the communication unit and host server by communicating the data via a sessionless-oriented communication protocol over a first communication channel between the virtual session manager and the communication unit, and by communicating the data via a session-oriented communication protocol between the virtual session manager and the host server.

8. A method of controlling communications with a communication unit (201) comprising:

at a communication server (220) in communication with the communication unit via a first communication channel,

receiving a first data unit, the first data unit being communicated between the communication unit and a further communication unit; characterized by

estimating a transmission value for the first data unit, and comparing the estimated transmission value with a transmission limit to determine whether to allow communication of further data units between the communication unit and the communication server.

9. The method of claim 8, wherein:

step (b) further comprises at least one of:

sending a notifier message to the communication unit when the transmission limit is less than a predetermined threshold value;

preventing communication of further data units via the communication server when the estimated transmission value exceeds the transmission limit;

determining an estimated total transmission value from the estimated transmission value and prior estimated transmission values for prior data units communicated between the communication unit and communication server, and sending a notifier message to the communication unit when the estimated total transmission value exceeds a predetermined threshold; and

determining an estimated total transmission value from the estimated transmission value and prior estimated transmission values for prior data units communicated between the communication unit and communication server, and preventing communication of further data units via the communication server when the estimated total transmission value exceeds the transmission limit.

10. A communication server characterized by a data transfer manager (229) operable for communicating data with a communication unit (201), the data transfer manager comprising a rate governor (234, 209) for estimating a transmission value for data communicated between communication unit and a further communication unit, and comparing the estimated transmission value with a transmission limit to determine whether to permit further communications with the communication unit via the communication server.

11. The communication server of claim 10, wherein the rate governor is further operable for at least one of:

preventing communication of further data between the communication server and communication unit when the estimated transmission value exceeds the transmission limit;

determining an estimated total transmission value from the estimated transmission value and prior estimated transmission values for prior data communicated between the communication unit and communication server, and preventing communication of further data units via the communication server when the estimated total transmission value exceeds the transmission limit; and

determining an estimated total transmission value from the estimated transmission value and prior estimated transmission values for prior data communicated between the communication unit and communication server, and sending a notifier message to the communication unit when the estimated total transmission value exceeds a predetermined threshold.

12. The communication server of claim 10, wherein the data transfer manager further comprises a virtual session manager adapted to control communication of data between the communication unit and a host server by communicating the data via a sessionless-oriented communication protocol over a first communication channel between the virtual session manager and the communication unit, and by communicating the data via a session-oriented communication protocol between the virtual session manager and the host server.

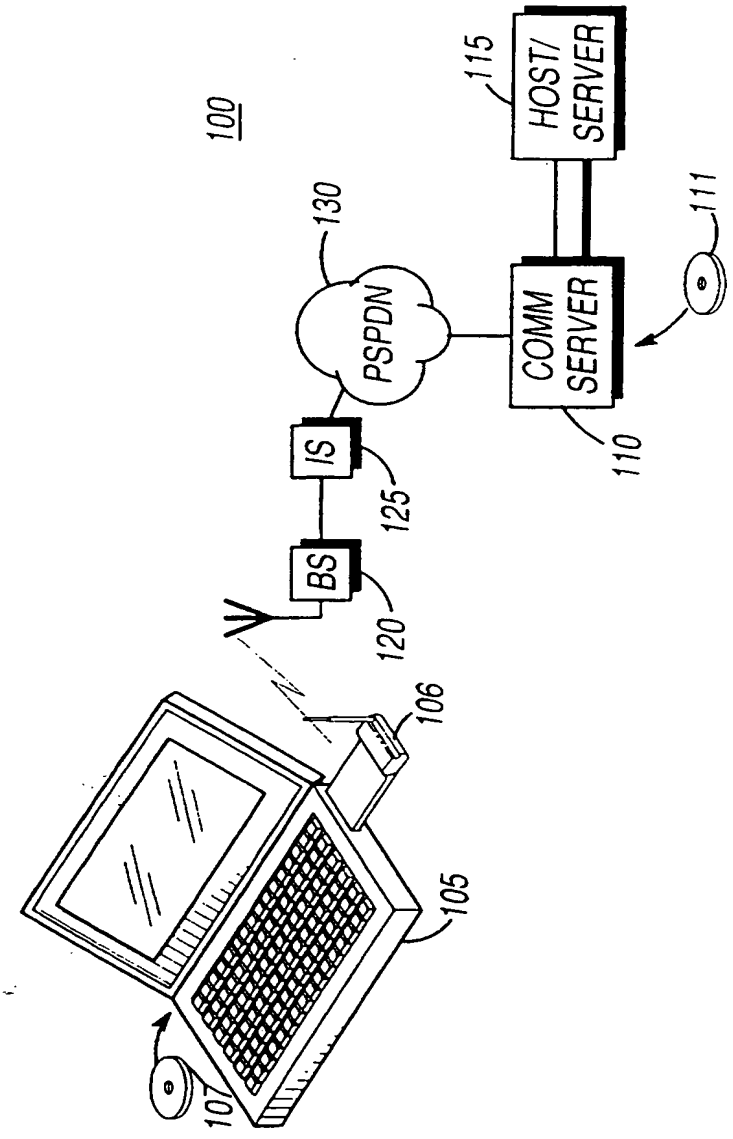

100
115
HOST/
SERVER
130
PSPDN
111
COMM
SERVER
110
IS
125
BS
120
106
105
107

FIG.1

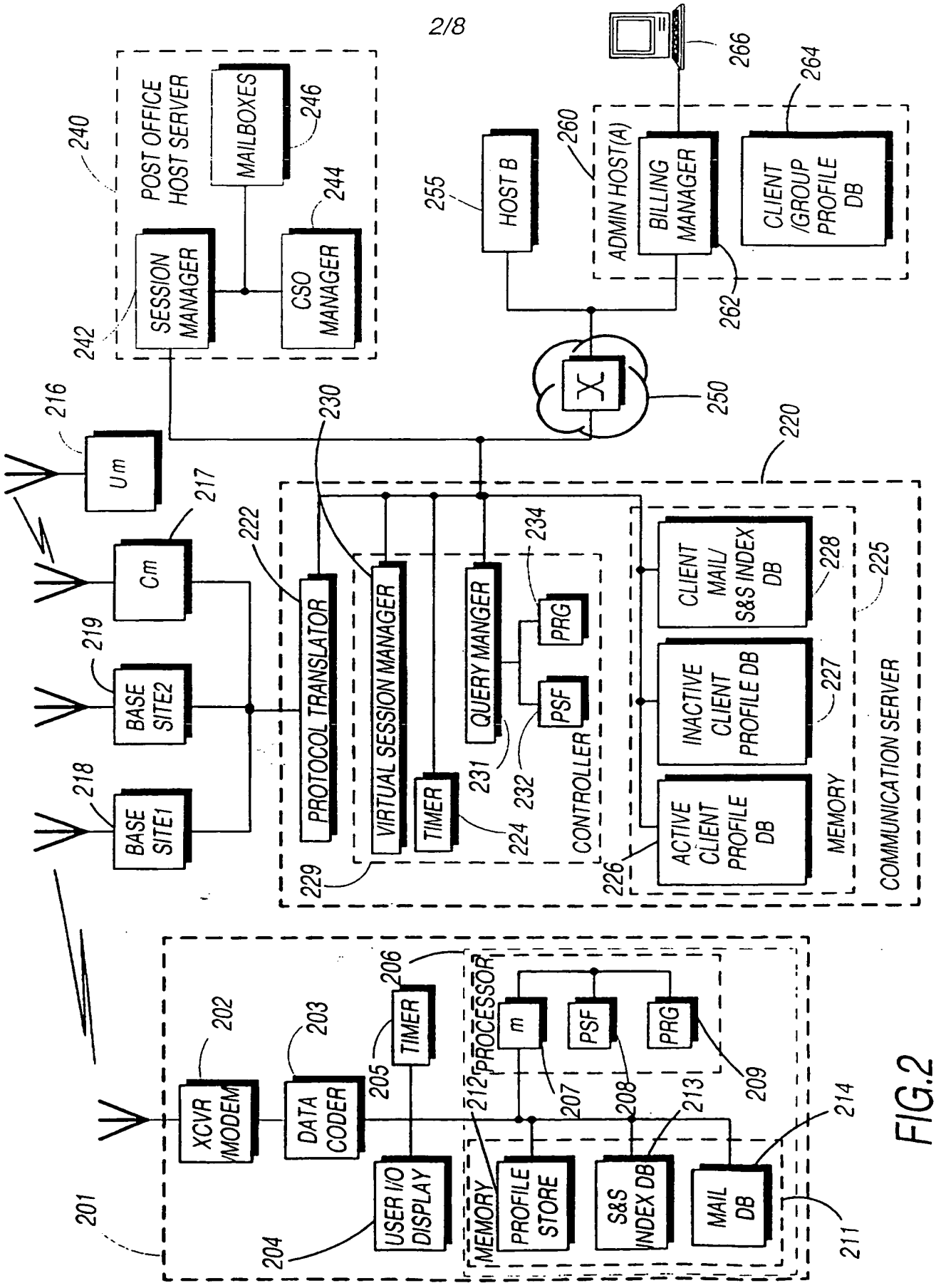

POST OFFICE HOST SERVER
240
MAILBOXES
246
SESSION MANAGER
242
CSO MANAGER
244
HOST B
255
ADMIN HOST(A)
260
BILLING MANAGER
262
CLIENT /GROUP PROFILE DB
264
266
X
250
Um
216
Cm
217
BASE SITE2
219
BASE SITE1
218
220
PROTOCOL TRANSLATOR
222
VIRTUAL SESSION MANAGER
229
TIMER
224
QUERY MANGER
231
PRG
234
PSF
232
CONTROLLER
230
CLIENT MAIL/ S&S INDEX DB
228
INACTIVE CLIENT PROFILE DB
227
ACTIVE CLIENT PROFILE DB
226
MEMORY
225
COMMUNICATION SERVER
201
XCVR /MODEM
202
DATA CODER
203
TIMER
205
206
USER I/O DISPLAY
204
PROCESSOR
212
m
207
PSF
208
PRG
209
213
MEMORY
PROFILE STORE
S&S INDEX DB
MAIL DB
214
211

FIG.2

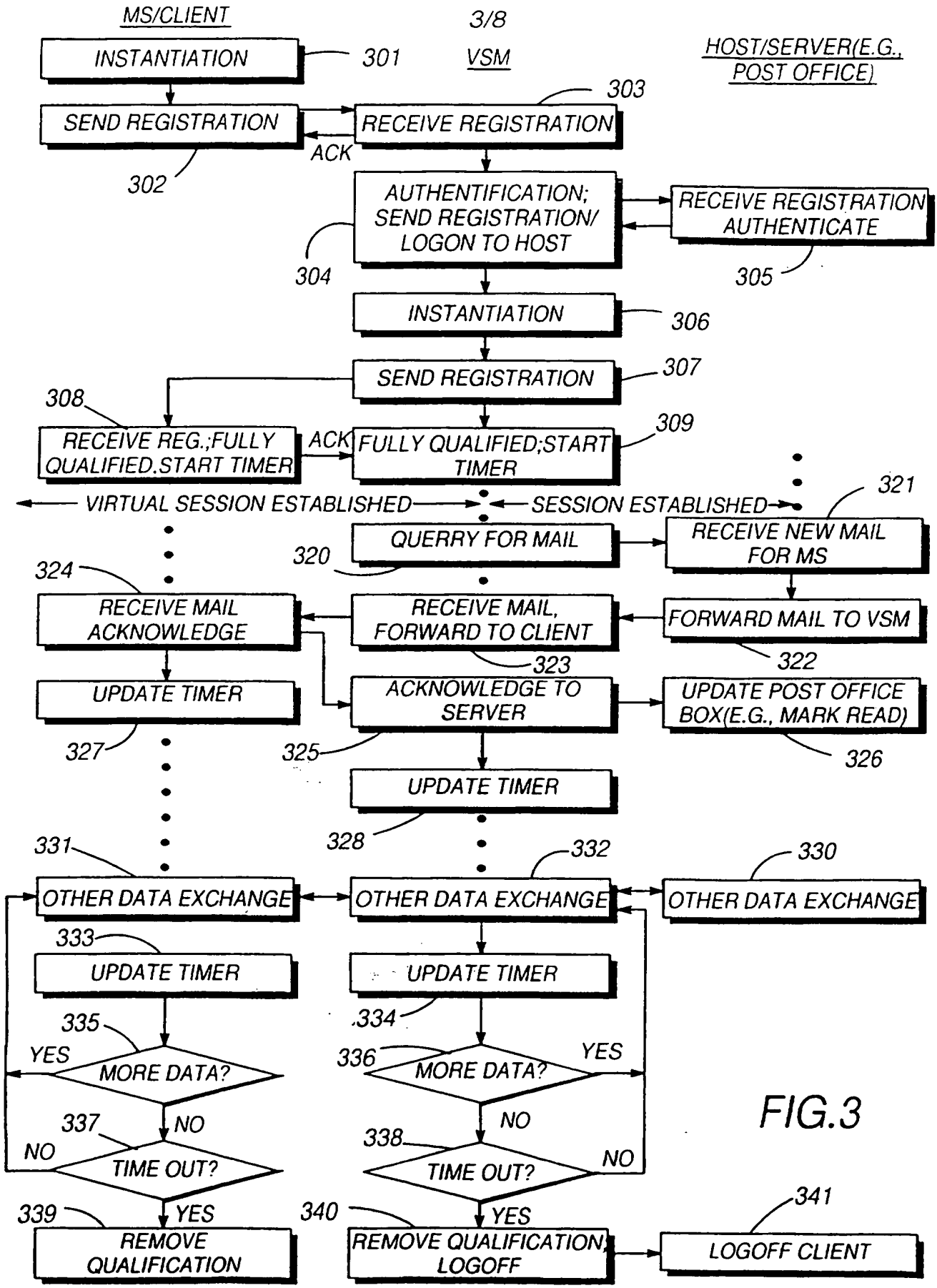

MS/CLIENT
3/8
VSM
HOST/SERVER(E.G., POST OFFICE)
INSTANTIATION
301
SEND REGISTRATION
RECEIVE REGISTRATION
303
ACK
302
AUTHENTIFICATION; SEND REGISTRATION/ LOGON TO HOST
RECEIVE REGISTRATION AUTHENTICATE
304
305
INSTANTIATION
306
SEND REGISTRATION
307
308
RECEIVE REG.;FULLY QUALIFIED.START TIMER
ACK
FULLY QUALIFIED;START TIMER
309
VIRTUAL SESSION ESTABLISHED
SESSION ESTABLISHED
321
QUERRY FOR MAIL
RECEIVE NEW MAIL FOR MS
320
324
RECEIVE MAIL ACKNOWLEDGE
RECEIVE MAIL, FORWARD TO CLIENT
FORWARD MAIL TO VSM
322
323
UPDATE TIMER
ACKNOWLEDGE TO SERVER
UPDATE POST OFFICE BOX(E.G., MARK READ)
327
325
326
UPDATE TIMER
328
331
332
330
OTHER DATA EXCHANGE
OTHER DATA EXCHANGE
OTHER DATA EXCHANGE
333
UPDATE TIMER
UPDATE TIMER
335
334
YES
MORE DATA?
336
MORE DATA?
YES
337
NO
NO
338
NO
TIME OUT?
TIME OUT?
339
YES
340
YES
341
REMOVE QUALIFICATION
REMOVE QUALIFICATION; LOGOFF
LOGOFF CLIENT

FIG.3

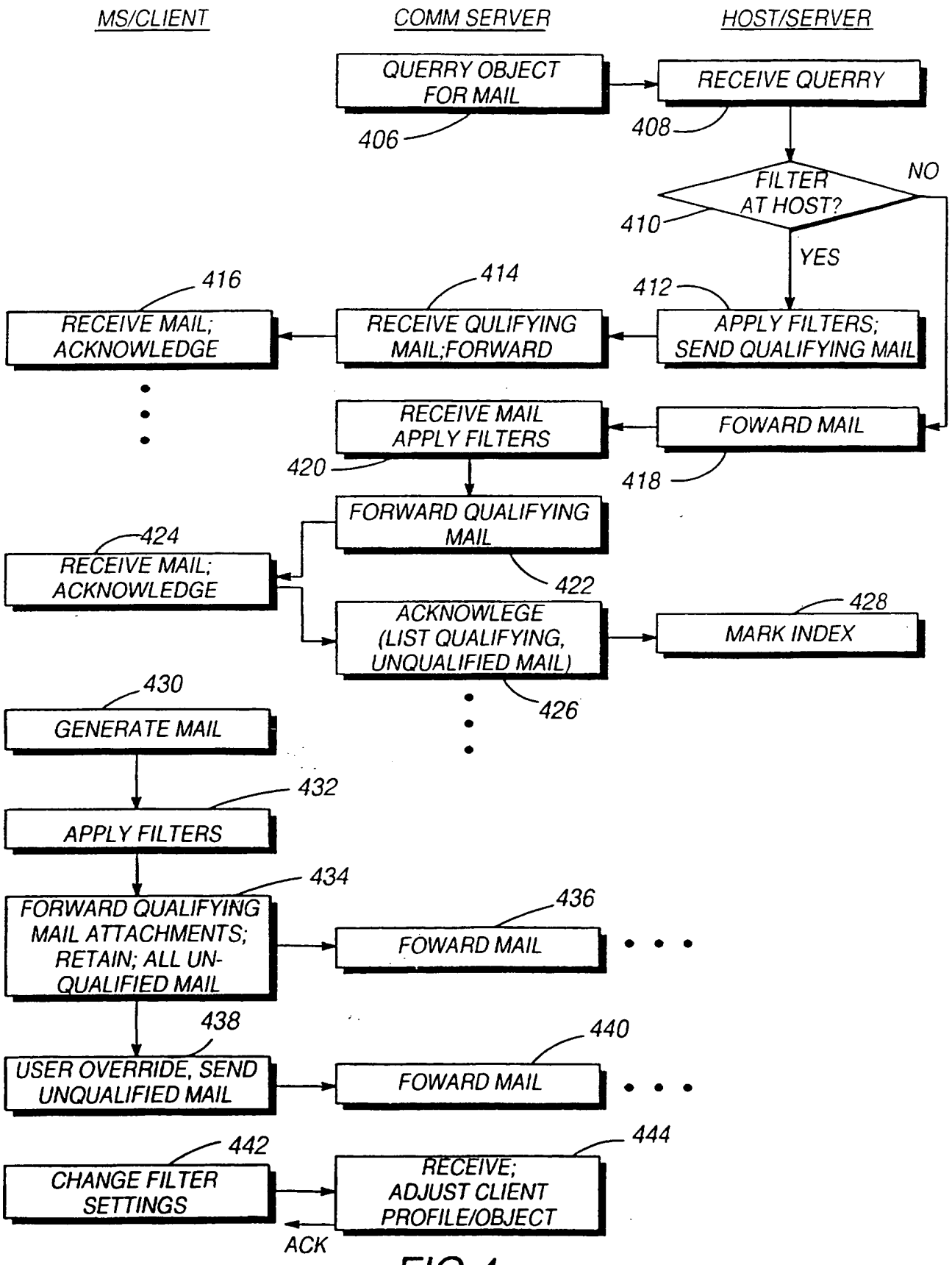

MS/CLIENT
COMM SERVER
HOST/SERVER
QUERRY OBJECT FOR MAIL
406
RECEIVE QUERRY
408
FILTER AT HOST?
410
NO
YES
APPLY FILTERS; SEND QUALIFYING MAIL
412
RECEIVE QULIFYING MAIL;FORWARD
414
RECEIVE MAIL; ACKNOWLEDGE
416
FOWARD MAIL
418
RECEIVE MAIL APPLY FILTERS
420
FORWARD QUALIFYING MAIL
422
RECEIVE MAIL; ACKNOWLEDGE
424
ACKNOWLEGE (LIST QUALIFYING, UNQUALIFIED MAIL)
426
MARK INDEX
428
GENERATE MAIL
430
APPLY FILTERS
432
FORWARD QUALIFYING MAIL ATTACHMENTS; RETAIN; ALL UN-QUALIFIED MAIL
434
FOWARD MAIL
436
USER OVERRIDE, SEND UNQUALIFIED MAIL
438
FOWARD MAIL
440
CHANGE FILTER SETTINGS
442
RECEIVE; ADJUST CLIENT PROFILE/OBJECT
444
ACK

FIG.4

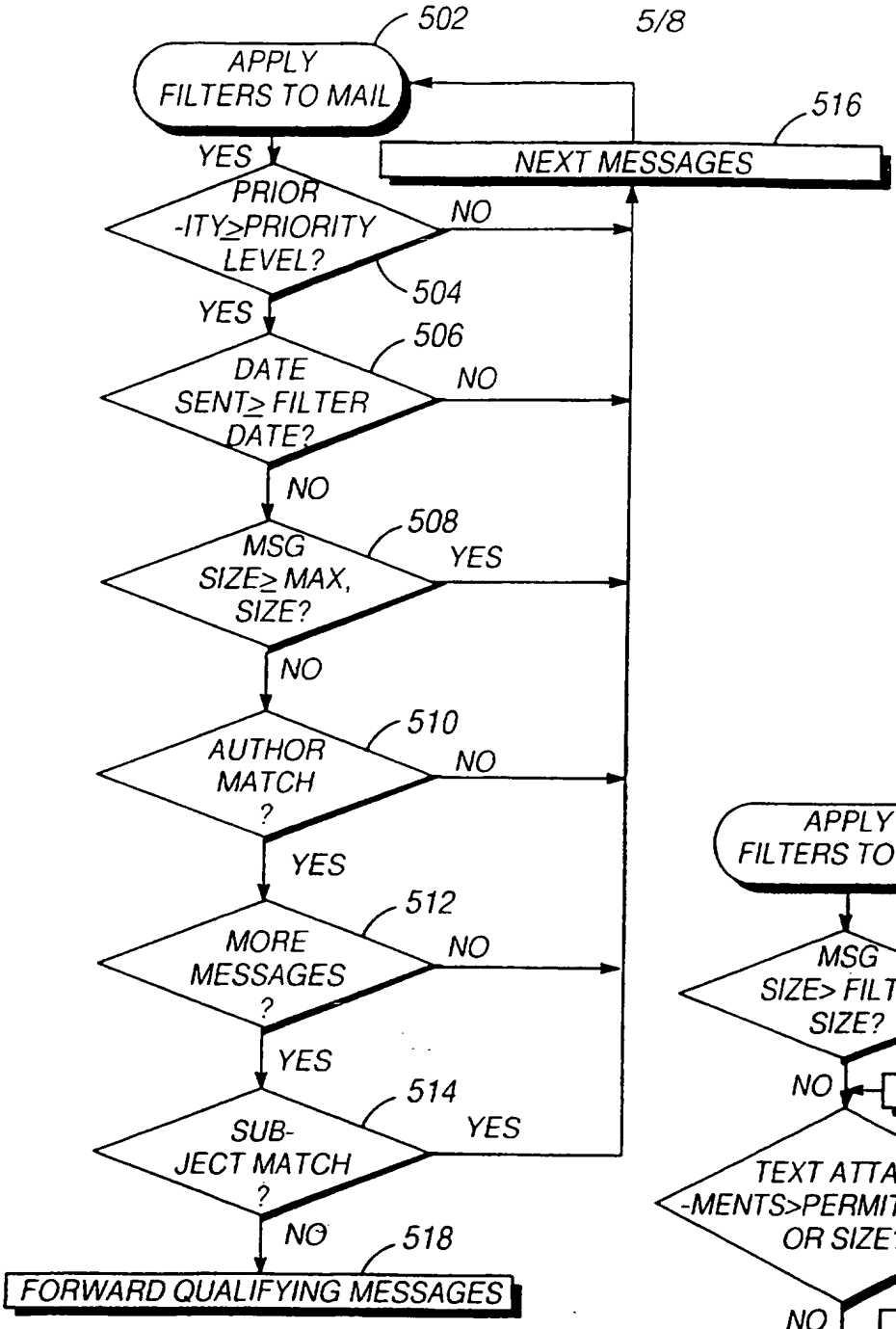

502
APPLY FILTERS TO MAIL
516
NEXT MESSAGES
YES
PRIOR -ITY≥PRIORITY LEVEL?
NO
504
YES
506
DATE SENT≥ FILTER DATE?
NO
NO
508
MSG SIZE≥ MAX, SIZE?
YES
NO
510
AUTHOR MATCH ?
NO
YES
512
MORE MESSAGES ?
NO
YES
514
SUB- JECT MATCH ?
YES
NO
518
FORWARD QUALIFYING MESSAGES

FIG.5

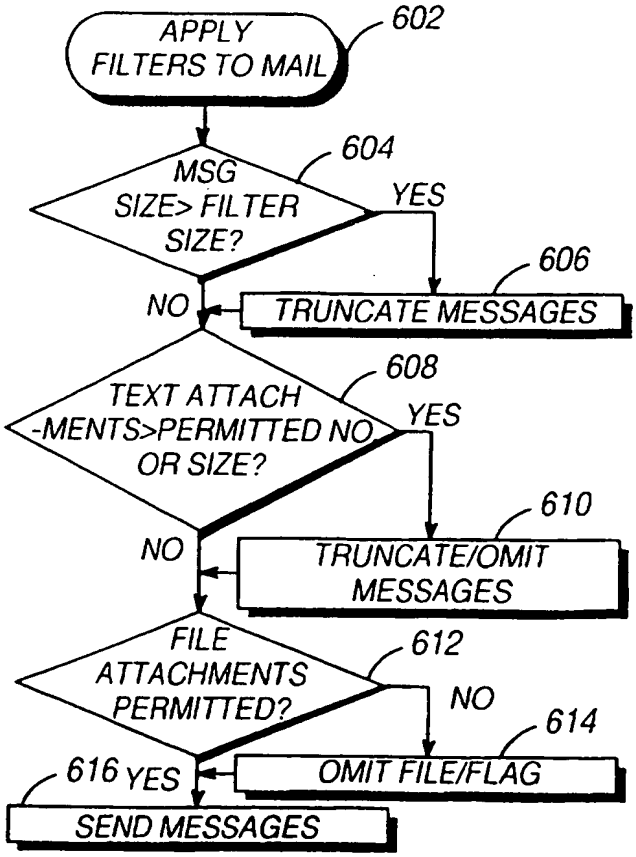

602
APPLY FILTERS TO MAIL
604
MSG SIZE> FILTER SIZE?
YES
606
NO
TRUNCATE MESSAGES
608
TEXT ATTACH -MENTS>PERMITTED NO OR SIZE?
YES
610
NO
TRUNCATE/OMIT MESSAGES
612
FILE ATTACHMENTS PERMITTED?
NO
614
616 YES
OMIT FILE/FLAG
SEND MESSAGES

FIG.6

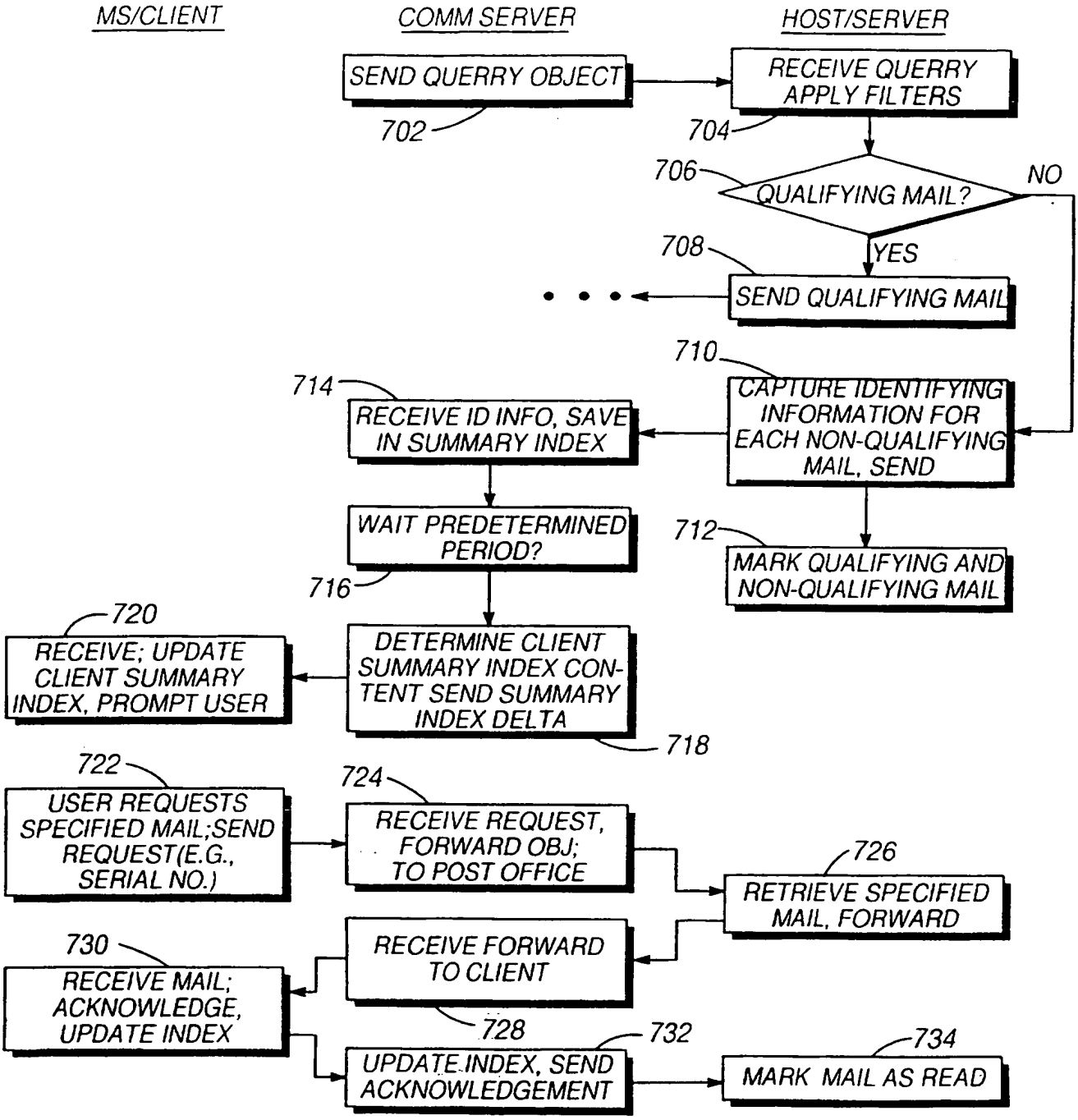


FIG.7

CLIENT 1 SUMMARY INDEX	801 SERIAL NO.1	802 HEADER INFO.1(E.G.,AUTHOR:SUBJECT,DATE/TIME IN; SIZE:ACKNOWLEDGEMENT/SIZE:PRIORITY)
	SERIAL NO.2	HEADER INFO 2
	⋮	⋮

FIG.8

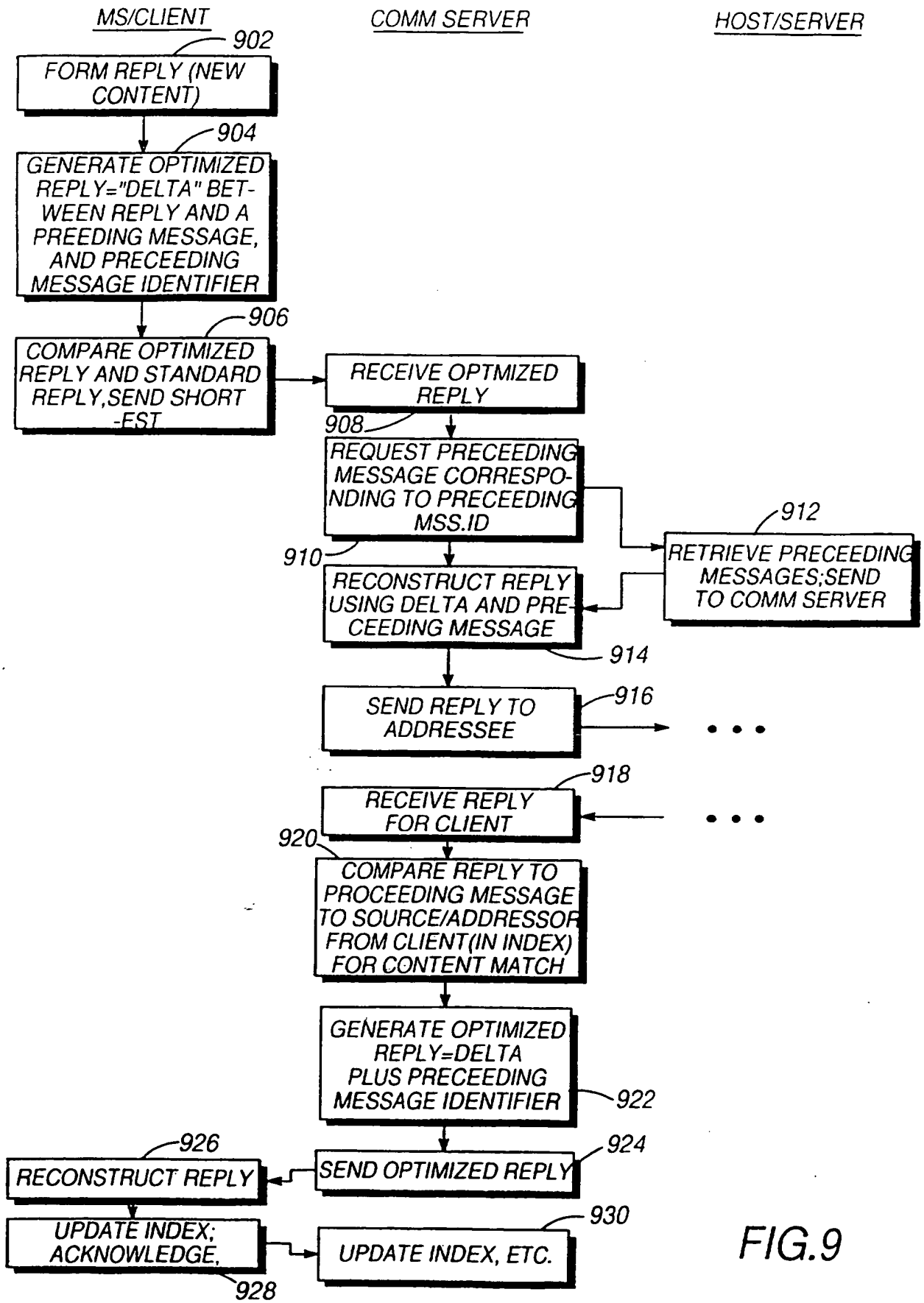

MS/CLIENT
COMM SERVER
HOST/SERVER
902
FORM REPLY (NEW CONTENT)
904
GENERATE OPTIMIZED REPLY="DELTA" BETWEEN REPLY AND A PREEDING MESSAGE, AND PRECEEDING MESSAGE IDENTIFIER
906
COMPARE OPTIMIZED REPLY AND STANDARD REPLY,SEND SHORT -EST
RECEIVE OPTMIZED REPLY
908
REQUEST PRECEEDING MESSAGE CORRESPONDING TO PRECEEDING MSS.ID
912
RETRIEVE PRECEEDING MESSAGES;SEND TO COMM SERVER
910
RECONSTRUCT REPLY USING DELTA AND PRECEEDING MESSAGE
914
SEND REPLY TO ADDRESSEE
916
RECEIVE REPLY FOR CLIENT
918
920
COMPARE REPLY TO PROCEEDING MESSAGE TO SOURCE/ADDRESSOR FROM CLIENT(IN INDEX) FOR CONTENT MATCH
GENERATE OPTIMIZED REPLY=DELTA PLUS PRECEEDING MESSAGE IDENTIFIER
922
SEND OPTIMIZED REPLY
924
RECONSTRUCT REPLY
926
UPDATE INDEX; ACKNOWLEDGE,
928
UPDATE INDEX, ETC.
930

FIG.9

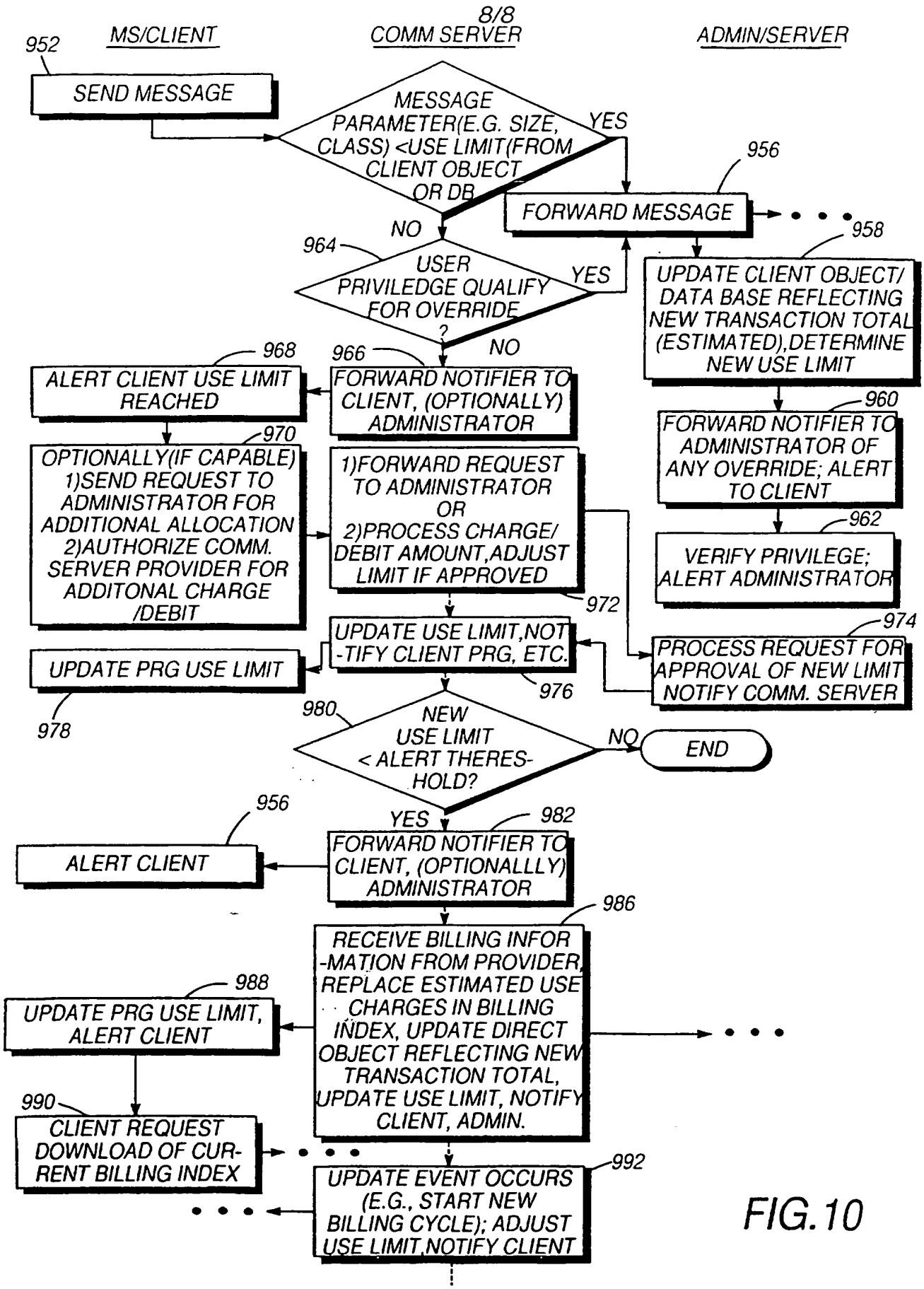

MS/CLIENT
COMM SERVER
ADMIN/SERVER
952
SEND MESSAGE
MESSAGE PARAMETER(E.G. SIZE, CLASS) <USE LIMIT(FROM CLIENT OBJECT OR DB
YES
NO
956
FORWARD MESSAGE
958
964
USER PRIVILEDGE QUALIFY FOR OVERRIDE ?
YES
NO
UPDATE CLIENT OBJECT/ DATA BASE REFLECTING NEW TRANSACTION TOTAL (ESTIMATED),DETERMINE NEW USE LIMIT
968
966
ALERT CLIENT USE LIMIT REACHED
FORWARD NOTIFIER TO CLIENT, (OPTIONALLY) ADMINISTRATOR
960
FORWARD NOTIFIER TO ADMINISTRATOR OF ANY OVERRIDE; ALERT TO CLIENT
970
OPTIONALLY(IF CAPABLE) 1)SEND REQUEST TO ADMINISTRATOR FOR ADDITIONAL ALLOCATION 2)AUTHORIZE COMM. SERVER PROVIDER FOR ADDITONAL CHARGE /DEBIT
1)FORWARD REQUEST TO ADMINISTRATOR OR 2)PROCESS CHARGE/ DEBIT AMOUNT,ADJUST LIMIT IF APPROVED
962
VERIFY PRIVILEGE; ALERT ADMINISTRATOR
972
974
UPDATE USE LIMIT,NOT -TIFY CLIENT PRG, ETC.
PROCESS REQUEST FOR APPROVAL OF NEW LIMIT NOTIFY COMM. SERVER
UPDATE PRG USE LIMIT
976
978
980
NEW USE LIMIT < ALERT THERES- HOLD?
NO
END
YES
956
982
ALERT CLIENT
FORWARD NOTIFIER TO CLIENT, (OPTIONALLLY) ADMINISTRATOR
986
RECEIVE BILLING INFOR -MATION FROM PROVIDER, REPLACE ESTIMATED USE CHARGES IN BILLING INDEX, UPDATE DIRECT OBJECT REFLECTING NEW TRANSACTION TOTAL, UPDATE USE LIMIT, NOTIFY CLIENT, ADMIN.
988
UPDATE PRG USE LIMIT, ALERT CLIENT
990
CLIENT REQUEST DOWNLOAD OF CUR- RENT BILLING INDEX
992
UPDATE EVENT OCCURS (E.G., START NEW BILLING CYCLE); ADJUST USE LIMIT,NOTIFY CLIENT

FIG. 10

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/19689

A. CLASSIFICATION OF SUBJECT MATTER
IPC(6) :G06F 13/00, 13/10
US CL :379/58, 59; 364/284.4, 284.3; 455/33.1; 395/200.09, 200.12
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 379/58, 59; 364/284.4, 284.3; 455/33.1; 395/200.09, 200.12

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
APS, MAYA

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5,109,401 A (HATTORI et al) 28 April 1992, Abstract; col. 5, line 38 - col. 6, line 47.	1-12
Y	US 5,138,650 A (STAHL et al) 11 August 1992, Abstract; col. 4, line 8 - col. 6 line 40.	1-12
Y	US 5,454,079 A (ROPER et al) 26 September 1995, col. 3, lines 22-44.	1-12
Y	US 5,287,456 A (RHODES et al) 15 February 1994, col. 1, line 18 - col. 2 line 46.	7 and 12

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be part of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search
26 FEBRUARY 1997

Date of mailing of the international search report
26 MAR 1997

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703) 305-3230

Authorized officer
THOMAS LEE
Telephone No. (703) 305-9717

Form PCT/ISA/210 (second sheet)(July 1992)*